





patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**Published:**

— *Without international search report and to be republished upon receipt of that report.*



5

10

**NOVEL HUMAN PROTEIN KINASES AND  
PROTEIN KINASE-LIKE ENZYMES**

The present invention claims priority on provisional application serial nos. 60/190,162; 60/174,185; 60/168,997; 60/179,364; 60/183,173; 60/178,078; 60/193,404; 60/195,953; and 60/187,150, all of which are hereby incorporated by reference in their entirety.

**FIELD OF THE INVENTION**

The present invention relates to kinase polypeptides, nucleotide sequences encoding the kinase polypeptides, as well as various products and methods useful for the diagnosis and treatment of various kinase-related diseases and conditions.

**BACKGROUND OF THE INVENTION**

The following description of the background of the invention is provided to aid in understanding the invention, but is not admitted to be or to describe prior art to the invention.

Cellular signal transduction is a fundamental mechanism whereby external stimuli that regulate diverse cellular processes are relayed to the interior of cells. One of the key biochemical mechanisms of signal transduction involves the reversible phosphorylation of proteins, which enables regulation of the activity of mature proteins by altering their structure and function.

Protein phosphorylation plays a pivotal role in cellular signal transduction. Among the biological functions controlled by this type of postranslational modification are: cell division, differentiation and death (apoptosis); cell motility and cytoskeletal structure; control of DNA

replication, transcription, splicing and translation; protein translocation events from the endoplasmic reticulum and Golgi apparatus to the membrane and extracellular space; protein nuclear import and export; regulation of metabolic reactions, etc. Abnormal protein phosphorylation is widely recognized to be causally linked to the etiology of many diseases  
5 including cancer as well as immunologic, neuronal and metabolic disorders.

The following abbreviations are used for kinases through this application:

	ASK	Apoptosis signal-regulating kinase
	CaMK	Ca <sup>2+</sup> /calmodulin-dependent protein kinase
	CCRK	Cell cycle-related kinase
10	CDK	Cyclin-dependent kinase
	CK	Casein kinase
	DAPK	Death-associated protein kinase
	DM	myotonic dystrophy kinase
	Dyrk	dual-specificity-tyrosine phosphorylating-regulated kinase
15	GAK	Cyclin G-associated kinase
	GRK	G-protein coupled receptor
	GuC	Guanylate cyclase
	HIPK	Homeodomain-interacting protein kinase
	IRAK	Interleukin-1 receptor-associated kinase
20	MAPK	Mitogen activated protein kinase
	MAST	Microtubule-associated STK
	MLCK	Myosin-light chain kinase
	MLK	Mixed lineage kinase
	NIMA	NimA-related protein kinase
25	PKA	cAMP-dependent protein kinase
	RSK	Ribosomal protein S6 kinase
	RTK	Receptor tyrosine kinase
	SGK	Serum and glucocorticoid-regulated kinase
	STK	serine threonine kinase
30	ULK	UNC-51-like kinase

The best-characterized protein kinases in eukaryotes phosphorylate proteins on the hydroxyl substituent of serine, threonine and tyrosine residues, which are the most common phospho-acceptor amino acid residues. However, phosphorylation on histidine has also been  
5 observed in bacteria.

The presence of a phosphate moiety modulates protein function in multiple ways. A common mechanism includes changes in the catalytic properties ( $V_{max}$  and  $K_m$ ) of an enzyme, leading to its activation or inactivation.

A second widely recognized mechanism involves promoting protein-protein interactions.  
10 An example of this is the tyrosine autophosphorylation of the ligand-activated EGF receptor tyrosine kinase. This event triggers the high-affinity binding to the phosphotyrosine residue on the receptor's C-terminal intracellular domain to the SH2 motif of the adaptor molecule Grb2. Grb2, in turn, binds through its SH3 motif to a second adaptor molecule, such as SHC. The formation of this ternary complex activates the signaling events that are responsible for the  
15 biological effects of EGF. Serine and threonine phosphorylation events also have been recently recognized to exert their biological function through protein-protein interaction events that are mediated by the high-affinity binding of phosphoserine and phosphothreonine to WW motifs present in a large variety of proteins (Lu, P.J. et al (1999) Science 283:1325-1328).

A third important outcome of protein phosphorylation is changes in the subcellular  
20 localization of the substrate. As an example, nuclear import and export events in a large diversity of proteins are regulated by protein phosphorylation (Drier E.A. et al (1999) Genes Dev 13: 556-568).

Protein kinases are one of the largest families of eukaryotic proteins with several hundred known members. These proteins share a 250-300 amino acid domain that can be subdivided into  
25 12 distinct subdomains that comprise the common catalytic core structure. These conserved protein motifs have recently been exploited using PCR-based and bioinformatic strategies leading to a significant expansion of the known kinases. Multiple alignment of the sequences in the catalytic domain of protein kinases and subsequent parsimony analysis permits their segregation into sub-families of related kinases.

Kinases largely fall into two groups: those specific for phosphorylating serines and threonines, and those specific for phosphorylating tyrosines. Some kinases, referred to as "dual specificity" kinases, are able to phosphorylate on tyrosine as well as serine/threonine residues.

Protein kinases can also be characterized by their location within the cell. Some kinases  
5 are transmembrane receptor-type proteins capable of directly altering their catalytic activity in response to the external environment such as the binding of a ligand. Others are non-receptor-type proteins lacking any transmembrane domain. They can be found in a variety of cellular compartments from the inner surface of the cell membrane to the nucleus.

Many kinases are involved in regulatory cascades wherein their substrates may include  
10 other kinases whose activities are regulated by their phosphorylation state. Ultimately the activity of some downstream effector is modulated by phosphorylation resulting from activation of such a pathway. The conserved protein motifs of these kinases have recently been exploited using PCR-based cloning strategies leading to a significant expansion of the known kinases.

Multiple alignment of the sequences in the catalytic domain of protein kinases and  
15 subsequent parsimony analysis permits the segregation of related kinases into distinct branches of subfamilies including: tyrosine kinases (PTK's), dual-specificity kinases, and serine/threonine kinases (STK's). The latter subfamily includes cyclic-nucleotide-dependent kinases, calcium/calmodulin kinases, cyclin-dependent kinases (CDK's), MAP-kinases, serine-threonine kinase receptors, and several other less defined subfamilies.

20 The protein kinases may be classified into several major groups including AGC, CAMK, Casein kinase 1, CMGC, STE, tyrosine kinases, and atypical kinases (Plowman, GD *et al.*, *Proceedings of the National Academy of Sciences*, USA, Vol. 96, Issue 24, 13603-13610, November 23, 1999; see also [www.kinase.com](http://www.kinase.com)). In addition, there are a number of minor yet distinct families, including families related to worm- or fungal-specific kinases, and a family  
25 designated "other" to represent several smaller families. Within each group are several distinct families of more closely related kinases. In addition, an "atypical" family represents those protein kinases whose catalytic domain has little or no primary sequence homology to conventional kinases, including the A6 kinases and PI3 kinases.

30 AGC group

The AGC kinases are basic amino acid-directed enzymes that phosphorylate residues found proximal to Arg and Lys. Examples of this group are the G protein-coupled receptor kinases (GRKs), the cyclic nucleotide-dependent kinases (PKA, PKC, PKG), NDR or DBF2  
5 kinases, ribosomal S6 kinases, AKT kinases, myotonic dystrophy kinases (DMPKs), MAPK interacting kinases (MNKs), MAST kinases, and Mo3C11.1\_ce family originally identified only in nematodes.

GRKs regulate signaling from heterotrimeric guanine protein coupled receptors (GPCRs). Mutations in GPCRs cause a number of human diseases, including retinitis pigmentosa,  
10 stationary night blindness, color blindness, hyperfunctioning thyroid adenomas, familial precocious puberty, familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism (OMIM, <http://www.ncbi.nlm.nih.gov/Omim/>). The regulation of GPCRs by GRKs indirectly implicates GRKs in these diseases.

The cAMP-dependent protein kinases (PKA) consist of heterotetramers comprised of 2  
15 catalytic (C) and 2 regulatory (R) subunits, in which the R subunits bind to the second messenger cAMP, leading to dissociation of the active C subunits from the complex. Many of these kinases respond to second messengers such as cAMP resulting in a wide range of cellular responses to hormones and neurotransmitters.

AKT is a mammalian proto-oncoprotein regulated by phosphatidylinositol 3-kinase (PI3-  
20 K), which appears to function as a cell survival signal to protect cells from apoptosis. Insulin receptor, RAS, PI3-K, and PDK1 all act as upstream activators of AKT, whereas the lipid phosphatase PTEN functions as a negative regulator of the PI3-K/AKT pathway. Downstream targets for AKT-mediated cell survival include the pro-apoptotic factors BAD and Caspase9, and transcription factors in the forkhead family, such as DAF-16 in the worm. AKT is also an  
25 essential mediator in insulin signaling, in part due to its use of GSK-3 as another downstream target.

The S6 kinases regulate a wide array of cellular processes involved in mitogenic response including protein synthesis, translation of specific mRNA species, and cell cycle progression from G1 to S phase. The gene has been localized to chromosomal region 17q23 and is amplified  
30 in breast cancer (Couch, *et al.*, Cancer Res. 1999 Apr 1;59(7):1408-11).

### CAMK Group

The CAMK kinases are also basic amino acid-directed kinases. They include the  
5 Ca<sup>2+</sup>/calmodulin-regulated and AMP-dependent protein kinases (AMPK), myosin light chain  
kinases (MLCK), MAP kinase activating protein kinases (MAPKAPKs) checkpoint 2 kinases  
(CHK2), death-associated protein kinases (DAPKs), phosphorylase kinase (PHK), Rac and Rho-  
binding Trio kinases, a “unique” family of CAMKs, and the EMK-related protein kinases.

The EMK family of STKs are involved in the control of cell polarity, microtubule  
10 stability and cancer. One member of the EMK family, C-TAK1, has been reported to control  
entry into mitosis by activating Cdc25C which in turn dephosphorylates Cdc2. Also included in  
the EMK family is MAKV, which has been shown to be overexpressed in metastatic tumors  
(*Dokl. Akad. Nauk* 354 (4), 554-556 (1997)).

### 15 CMGC Group

The CMGC kinases are “proline-directed” enzymes phosphorylating residues that exist in  
a proline-rich context. They include the cyclin-dependent kinases (CDKs), mitogen-activated  
protein kinases (MAPKs), GSK3s, RCKs, and CLKs. Most CMGC kinases have larger-than-  
20 average kinase domains owing to the presence of insertions within subdomains X and XI.

CDK's play a pivotal role in the regulation of mitosis during cell division. The process of  
cell division occurs in four stages: S phase, the period during which chromosomes duplicate, G2,  
mitosis and G1 or interphase. During mitosis the duplicated chromosomes are evenly segregated  
allowing each daughter cell to receive a complete copy of the genome. A key mitotic regulator  
25 in all eukaryotic cells is the STK cdc2, a CDK regulated by cyclin B. However some CDK-like  
kinases, such as CDK5 are not cyclin associated nor are they cell cycle regulated.

MAPKs play a pivotal role in many cellular signaling pathways, including stress response  
and mitogenesis (Lewis, T. S., Shapiro, P. S., and Ahn, N. G. (1998) *Adv. Cancer Res.* 74, 49-  
139). MAP kinases can be activated by growth factors such as EGF, and cytokines such as TNF-  
30 alpha. In response to EGF, Ras becomes activated and recruits Raf1 to the membrane where

Raf1 is activated by mechanisms that may involve phosphorylation and conformational changes (Morrison, D. K., and Cutler, R. E. (1997) *Curr. Opin. Cell Biol.* 9, 174-179). Active Raf1 phosphorylates MEK1 which in turn phosphorylates and activates the ERKs.

5        Tyrosine Protein Kinase Group

The tyrosine kinase group encompass both cytoplasmic (*e.g. src*) as well as transmembrane receptor tyrosine kinases (*e.g.* EGF receptor). These kinases play a pivotal role in the signal transduction processes that mediate cell proliferation, differentiation and apoptosis.

10    One of the sequences, 17000030181412, is related to the human RET kinase. Mutations of the RET gene, encoding a receptor tyrosine kinase, have been associated with the inherited cancer syndromes MEN 2A and MEN 2B. They have also further been associated with both familial and sporadic medullary thyroid carcinomas. The kinase activity can be aberrantly activated by missense mutations affecting cysteine residues within the extracellular domain, leading to potent

15    oncogenicity (*Oncogene* 1999 Aug 26;18(34):4833-8).

STE Group

The STE family refers to the 3 classes of protein kinases that lie sequentially upstream of

20    the MAPKs. This group includes STE7 (MEK or MAPKK) kinases, STE11 (MEKK or MAPKKK) kinases and STE20 (MEKKK) kinases. In humans, several protein kinase families that bear only distant homology with the STE11 family also operate at the level of MAPKKKs including RAF, MLK, TAK1, and COT. Since crosstalk takes place between protein kinases functioning at different levels of the MAPK cascade, the large number of STE family kinases

25    could translate into an enormous potential for upstream signal specificity.

The prototype STE20 from baker's yeast is regulated by a hormone receptor, signaling to directly affect cell cycle progression through modulation of CDK activity. It also coordinately regulates changes in the cytoskeleton and in transcriptional programs in a bifurcating pathway. In a similar way, the homologous kinases in humans are likely to play a role in extracellular

30    regulation of growth, cell adhesion and migration, and changes in transcriptional programs, all

three of which have critical roles in tumorigenesis. Mammalian STE20-related protein kinases have been implicated in response to growth factors or cytokines, oxidative-, UV-, or irradiation-related stress pathways, inflammatory signals (e.g.  $\text{TNF}\alpha$ ), apoptotic stimuli (e.g. Fas), T and B cell costimulation, the control of cytoskeletal architecture, and cellular transformation.

- 5 Typically the STE20-related kinases serve as upstream regulators of MAPK cascades. Examples include: HPK1, a protein-serine/threonine kinase (STK) that possesses a STE20-like kinase domain that activates a protein kinase pathway leading to the stress-activated protein kinase SAPK/JNK; PAK1, an STK with an upstream CDC42-binding domain that interacts with Rac and plays a role in cellular transformation through the Ras-MAPK pathway; and murine NIK,  
10 which interacts with upstream receptor tyrosine kinases and connects with downstream STE11-family kinases.

- NEK kinases are related to NIMA, which is required for entry into mitosis in the filamentous fungus *A. nidulans*. Mutations in the *nimA* gene cause the *nim* (never in mitosis) G2 arrest phenotype in this fungus (Fry, A.M. and Nigg, E.A. (1995) *Current Biology* 5: 1122-  
15 1125). Several observations suggest that higher eukaryotes may have a NIMA functional counterpart(s): (1) expression of a dominant-negative form of NIMA in HeLa cells causes a G2 arrest; (2) overexpression of NIMA causes chromatin condensation, not only in *A. nidulans*, but also in yeast, *Xenopus* oocytes and HeLa cells (Lu, K.P. and Hunter, T. (1995) *Prog. Cell Cycle Res.* 1, 187-205); (3) NIMA when expressed in mammalian cells interacts with *pin1*, a prolyl-  
20 prolyl isomerase that functions in cell cycle regulation (Lu, K.P. *et al.* (1996) *Nature* 380, 544-547); (4) okadaic acid inhibitor studies suggests the presence of *cdc2*-independent mechanism to induce mitosis (Ghosh, S. *et al.* (1998) *Exp. Cell Res.* 242, 1-9) and (5) a NIMA-like kinase (*fin1*) exists in another eukaryote besides *Aspergillus*, *Saccharomyces pombe* (Krien, M.J.E. *et al.* (1998) *J. Cell Sci.* 111, 967-976). Four mammalian NIMA-like kinases have been identified.  
25 NEK1, NEK2, NEK3 and NRK2. Despite the similarity of the NIMA-related kinases to NIMA over the catalytic region, the mammalian kinases are structurally different to NIMA over the extracatalytic regions. In addition the mammalian kinases are unable to complement the *nim* phenotype in *Aspergillus nimA* mutants. These observations lead to the following three possibilities: 1) the mammalian NIMA homologue remains unidentified; 2) there is no NIMA  
30 homologue in higher eukaryotes; 3) the biological function of NIMA is carried out by multiple,



related kinases in higher eukaryotes. The elucidation and biological characterization of additional mammalian NIMA- and NEK-related kinases should assist in elucidating this question.

### Casein Kinase 1 Group

5

The CK1 family represents a distant branch of the protein kinase family. The hallmarks of protein kinase subdomains VIII and IX are difficult to identify. One or more forms are ubiquitously distributed in mammalian tissues and cell lines. CK1 kinases are found in cytoplasm, in nuclei, membrane-bound, and associated with the cytoskeleton. Splice variants  
10 differ in their subcellular distribution.

### "Other" Group

Several families cluster within a group of unrelated kinases termed "Other". Included  
15 are: CHK1; Elongation 2 factor kinases (EIFK); homologues of the yeast sterile family kinases (STE), which refers to 3 classes of kinases which lie sequentially upstream of the MAPKs; Calcium-calmodulin kinase kinases (CAMKK); dual-specific tyrosine kinases (DYRK); I $\kappa$ B kinases (IKK); Integrin receptor kinase (IRAK); endoribonuclease-associated kinases (IRE); Mixed lineage kinase (MLK); LIM-domain containing kinase (LIMK); MOS; PIM; Receptor  
20 interacting kinase (RIP); SR-protein specific kinase (SRPK); RAF; Serine-threonine kinase receptors (STKR); TAK1; Testis specific kinase (TSK); tousel-related kinase (TSL); UNC51-related kinase (UNC); VRK; WEE; mitotic kinases (BUB1, AURORA, PLK, and NIMA/NEK); several families that are close homologues to worm (C26C2.1, YQ09, ZC581.9, YFL033c, C24A1.3); Drosophila (SLOB), or yeast (YDOD\_sp, YGR262\_sc) kinases; and others that are  
25 "unique," that is, those which do not cluster into any obvious family. Additional families are even less well defined and first were identified in lower eukaryotes such as yeast or worms (YNL020, YPL236, YQ09, YWY3, SCY1, C01H6.9, C26C2.1)

RIP2 is a serine-threonine kinase associated with the tumor necrosis factor (TNF) receptor complex and is implicated in the activation of NF-kappa B and cell death in mammalian  
30 cells. It has recently been demonstrated that RIP2 activates the MAPK pathway (Navas, *et al.*, *J*

*Biol. Chem.* 1999 Nov 19;274(47):33684-33690). RIP2 activates AP-1 and serum response element regulated expression by inducing the activation of the Elk1 transcription factor. RIP2 directly phosphorylates and activates ERK2 *in vivo* and *in vitro*. RIP2 in turn is activated through its interaction with Ras-activated Raf1. These results highlight the integrated nature of kinase signaling pathway.

The tousled (TSL) kinase was first identified in the plant *Arabidopsis thaliana*. TSL encodes a serine/threonine kinase that is essential for proper flower development. Human tousled-like kinases (Tlks) are cell-cycle-regulated enzymes, displaying maximal activities during S phase. This regulated activity suggests that Tlk function is linked to ongoing DNA replication (Sillje, *et al.*, *EMBO J* 1999 Oct 15;18(20):5691-5702).

#### Atypical Protein Kinase Group

There are several proteins with protein kinase activity that appear structurally unrelated to the eukaryotic protein kinases. These include; *Dictyostelium* myosin heavy chain kinase A (MHCKA), *Physarum polycephalum* actin-fragmin kinase, the human A6 PTK, human BCR, mitochondrial pyruvate dehydrogenase and branched chain fatty acid dehydrogenase kinase, and the prokaryotic "histidine" protein kinase family. The slime mold, worm, and human eEF-2 kinase homologues have all been demonstrated to have protein kinase activity, yet they bear little resemblance to conventional protein kinases except for the presence of a putative GxGxxG ATP-binding motif.

The so-called histidine kinases are abundant in prokaryotes, with more than 20 representatives in *E. coli*, and have also been identified in yeast, molds, and plants. In response to external stimuli, these kinases act as part of two-component systems to regulate DNA replication, cell division, and differentiation through phosphorylation of an aspartate in the target protein. To date, no "histidine" kinases have been identified in metazoans, although mitochondrial pyruvate dehydrogenase (PDK) and branched chain alpha-ketoacid dehydrogenase kinase (BCKD kinase), are related in sequence. PDK and BCKD kinase represent a unique family of atypical protein kinases involved in regulation of glycolysis, the citric acid cycle, and protein synthesis during protein malnutrition. Structurally they conserve only the C-terminal

portion of "histidine" kinases including the G box regions. BCKD kinase phosphorylates the E1a subunit of the BCKD complex on Ser-293, proving it to be a functional protein kinase. Although no bona fide "histidine" kinase has yet been identified in humans, they do contain PDK.

Several other proteins contain protein kinase-like homology including: receptor guanylyl  
5 cyclases, diacylglycerol kinases, choline/ethanolamine kinases, and YLK1-related antibiotic resistance kinases. Each of these families contain short motifs that were recognized by our profile searches with low scoring E-values, but *a priori* would not be expected to function as protein kinases. Instead, the similarity could simply reflect the modular nature of protein evolution and the primal role of ATP binding in diverse phosphotransfer enzymes. However, two  
10 recent papers on a bacterial homologue of the YLK1 family suggests that the aminoglycoside phosphotransferases (APHs) are structurally and functionally related to protein kinases. There are over 40 APHs identified from bacteria that are resistant to aminoglycosides such as kanamycin, gentamycin, or amikacin. The crystal structure of one well characterized APH reveals that it shares greater than 40% structural identity with the 2 lobed structure of the  
15 catalytic domain of cAMP-dependent protein kinase (PKA), including an N-terminal lobe composed of a 5-stranded antiparallel beta sheet and the core of the C-terminal lobe including several invariant segments found in all protein kinases. APHs lack the GxGxxG normally present in the loop between beta strands 1 and 2 but contain 7 of the 12 strictly conserved residues present in most protein kinases, including the HGDxxxN signature sequence in kinase  
20 subdomain VIB. Furthermore, APH also has been shown to exhibit protein-serine/threonine kinase activity, suggesting that other YLK-related molecules may indeed be functional protein kinases.

The eukaryotic lipid kinases (PI3Ks, PI4Ks, and PIPKs) also contain several short motifs similar to protein kinases, but otherwise share minimal primary sequence similarity. However,  
25 once again structural analysis of PIPKII-beta defines a conserved ATP-binding core that is strikingly similar to conventional protein kinases. Three residues are conserved among all of these enzymes including (relative to the PKA sequence) Lys-72 which binds the gamma-phosphate of ATP, Asp-166 which is part of the HRDLK motif and Asp-184 from the conserved  $Mg^{++}$  or  $Mn^{++}$  binding DFG motif. The worm genome contains 12 phosphatidylinositol kinases,  
30 including 3 PI3-kinases, 2 PI4-kinases, 3 PIP5-kinases, and 4 PI3-kinase-related kinases. The

latter group has 4 mammalian members (DNA-PK, FRAP/TOR, ATM, and ATR), which have been shown to participate in the maintenance of genomic integrity in response to DNA damage, and exhibit true protein kinase activity, raising the possibility that other PI-kinases may also act as protein kinases. Regardless of whether they have true protein kinase activity, PI3-kinases are  
5 tightly linked to protein kinase signaling, as evidenced by their involvement downstream of many growth factor receptors and as upstream activators of the cell survival response mediated by the AKT protein kinase.

### SUMMARY OF THE INVENTION

The present invention relates, in part, to human protein kinases and protein kinase-like enzymes identified from genomic sequencing.

5 Tyrosine and serine/threonine kinases (PTK's and STK's) have been identified and their protein sequence predicted as part of the instant invention. Mammalian members of these families were identified through the use of a bioinformatics strategy. The partial or complete sequences of these kinases are presented here, together with their classification, predicted or deduced protein structure.

10 One aspect of the invention features an identified, isolated, enriched, or purified nucleic acid molecule encoding a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID  
15 NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ  
20 ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114.

The term "identified" in reference to a nucleic acid is meant that a sequence was selected from a genomic, EST, or cDNA sequence database based on it being predicted to encode a  
25 portion of a previously unknown or novel protein kinase.

By "isolated," in reference to nucleic acid, is meant a polymer of 10 (preferably 21, more preferably 39, most preferably 75) or more nucleotides conjugated to each other, including DNA and RNA that is isolated from a natural source or that is synthesized as the sense or complementary antisense strand. In certain embodiments of the invention, longer nucleic acids  
30 are preferred, for example those of 300, 600, 900, 1200, 1500, or more nucleotides and/or those having at least 50%, 60%, 75%, 80%, 85%, 90%, 95% or 99% identity to a sequence selected

from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, 5 SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, 10 SEQ ID NO:56, and SEQ ID NO:57.

The isolated nucleic acid of the present invention is unique in the sense that it is not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular (*i.e.*, chromosomal) environment. 15 Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide chain present, but that it is essentially free (about 90 - 95% pure at least) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

By the use of the term "enriched" in reference to nucleic acid is meant that the specific 20 DNA or RNA sequence constitutes a significantly higher fraction (2- to 5-fold) of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two. However, it should be noted 25 that enriched does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term "significant" is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other nucleic acids of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no 30 DNA or RNA from other sources. The DNA from other sources may, for example, comprise

DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes from naturally occurring events, such as viral infection, or tumor-type growths, in which the level of one mRNA may be naturally increased relative to other species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to  
5 elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation). Instead, it represents an indication that the sequence is relatively more pure than in the natural environment (compared to the natural level this level should be at  
10 least 2- to 5-fold greater, *e.g.*, in terms of mg/mL). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA  
15 library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately  $10^6$ -fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three  
20 orders, and more preferably four or five orders of magnitude is expressly contemplated.

By a "kinase polypeptide" is meant 32 (preferably 40, more preferably 45, most preferably 55) or more contiguous amino acids in a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ  
25 ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID  
30 NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ

ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. In certain aspects, polypeptides of 100, 200, 300, 400, 450, 500, 550, 600, 700, 800, 900 or more amino acids are preferred. The kinase polypeptide can be encoded by a full-length nucleic acid sequence or any portion (*e.g.*, a "fragment" as defined herein) of the full-length nucleic acid sequence, so long as a functional activity of the polypeptide is retained, including, for example, a catalytic domain, as defined herein, or a portion thereof. One of skill in the art would be able to select those catalytic domains, or portions thereof, which exhibit a kinase or kinase-like activity, *e.g.*, catalytic activity, as defined herein. It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. Equally, it is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or polypeptide which retains the functionality of the original. Such substitutions may include the replacement of an amino acid by a residue having similar physicochemical properties, such as substituting one aliphatic residue (Ile, Val, Leu or Ala) for another, or substitution between basic residues Lys and Arg, acidic residues Glu and Asp, amide residues Gln and Asn, hydroxyl residues Ser and Tyr, or aromatic residues Phe and Tyr. Further information regarding making amino acid exchanges which have only slight, if any, effects on the overall protein can be found in Bowie *et al.*, *Science*, 1990, 247, 1306-1310, which is incorporated herein by reference in its entirety including any figures, tables, or drawings. In all cases, all permutations are intended to be covered by this disclosure.

The amino acid sequence of a kinase peptide of the invention will be substantially similar to a sequence having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ



ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, or the corresponding full-length amino acid sequence, or fragments thereof.

A sequence that is substantially similar to a sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, will preferably have at least 90% identity (more preferably at least 95% and most preferably 99-100%) to the sequence.

By "identity" is meant a property of sequences that measures their similarity or relationship. Identity is measured by dividing the number of identical residues by the total number of residues and gaps and multiplying the product by 100. "Gaps" are spaces in an alignment that are the result of additions or deletions of amino acids. Thus, two copies of exactly the same sequence have 100% identity, but sequences that are less highly conserved, and have deletions, additions, or replacements, may have a lower degree of identity. Those skilled in the art will recognize that several computer programs are available for determining sequence identity using standard parameters, for example Gapped BLAST or PSI-BLAST (Altschul, *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402), BLAST (Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-410), and Smith-Waterman (Smith, *et al.* (1981) *J. Mol. Biol.* 147:195-197). Preferably, the default settings of these programs will be employed, but those skilled in the art recognize whether these settings need to be changed and know how to make the changes.

"Similarity" is measured by dividing the number of identical residues plus the number of conservatively substituted residues (see Bowie, *et al. Science*, 1999), 247, 1306-1310, which is incorporated herein by reference in its entirety, including any drawings, figures, or tables) by the total number of residues and gaps and multiplying the product by 100.

- 5 In preferred embodiments, the invention features isolated, enriched, or purified nucleic acid molecules encoding a kinase polypeptide comprising a nucleotide sequence that: (a) encodes a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114; (b) is the complement of the nucleotide sequence of (a); (c) hybridizes under highly stringent conditions to the nucleotide molecule of (a) and encodes a naturally occurring kinase polypeptide; (d) encodes a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ

ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, except that it lacks one or more, but not all, of the domains selected from the group consisting of an N-terminal domain, a catalytic domain, a C-terminal catalytic domain, a C-terminal domain, a coiled-coil structure region, a proline-rich region, a spacer region, and a C-terminal tail; and (e) is the complement of the nucleotide sequence of (d).

The term "complement" refers to two nucleotides that can form multiple favorable interactions with one another. For example, adenine is complementary to thymine as they can form two hydrogen bonds. Similarly, guanine and cytosine are complementary since they can form three hydrogen bonds. A nucleotide sequence is the complement of another nucleotide sequence if all of the nucleotides of the first sequence are complementary to all of the nucleotides of the second sequence.

Various low or high stringency hybridization conditions may be used depending upon the specificity and selectivity desired. These conditions are well known to those skilled in the art. Under stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having more than 1 or 2 mismatches out of 20 contiguous nucleotides, more preferably, such conditions prevent hybridization of nucleic acids having more than 1 or 2 mismatches out of 50 contiguous nucleotides, most preferably, such conditions prevent hybridization of nucleic acids having more than 1 or 2 mismatches out of 100 contiguous nucleotides. In some instances, the conditions may prevent hybridization of nucleic acids having more than 5 mismatches in the full-length sequence.

By stringent hybridization assay conditions is meant hybridization assay conditions at least as stringent as the following: hybridization in 50% formamide, 5X SSC, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 6.8, 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5X Denhardt's solution at 42 °C overnight; washing with 2X SSC, 0.1% SDS at 45 °C; and washing with 0.2X SSC, 0.1% SDS at 45 °C. Under some of the most stringent hybridization assay conditions, the second wash can be done with 0.1X SSC at a temperature up to 70 °C (Berger *et al.* (1987) Guide to Molecular Cloning Techniques pg 421, hereby incorporated by reference herein in its entirety including any figures, tables, or drawings.). However, other applications may require the use of conditions falling between these sets of conditions. Methods of determining the conditions

required to achieve desired hybridizations are well known to those with ordinary skill in the art, and are based on several factors, including but not limited to, the sequences to be hybridized and the samples to be tested. Washing conditions of lower stringency frequently utilize a lower temperature during the washing steps, such as 65 °C, 60 °C, 55 °C, 50 °C, or 42 °C.

5       The term "domain" refers to a region of a polypeptide which serves a particular function. For instance, N-terminal or C-terminal domains of signal transduction proteins can serve functions including, but not limited to, binding molecules that localize the signal transduction molecule to different regions of the cell or binding other signaling molecules directly responsible for propagating a particular cellular signal. Some domains can be expressed separately from the  
10 rest of the protein and function by themselves, while others must remain part of the intact protein to retain function. The latter are termed functional regions of proteins and also relate to domains.

      The term "N-terminal domain" refers to the extracatalytic region located between the initiator methionine and the catalytic domain of the protein kinase. The N-terminal domain can be identified following a Smith-Waterman alignment of the protein sequence against the non-  
15 redundant protein database to define the N-terminal boundary of the catalytic domain. Depending on its length, the N-terminal domain may or may not play a regulatory role in kinase function. An example of a protein kinase whose N-terminal domain has been shown to play a regulatory role is PAK65, which contains a CRIB motif used for Cdc42 and rac binding (Burbelo, P.D. *et al.* (1995) *J. Biol. Chem.* 270, 29071-29074).

20       The term "catalytic domain" refers to a region of the protein kinase that is typically 25-300 amino acids long and is responsible for carrying out the phosphate transfer reaction from a high-energy phosphate donor molecule such as ATP or GTP to itself (autophosphorylation) or to other proteins (exogenous phosphorylation). The catalytic domain of protein kinases is made up of 12 subdomains that contain highly conserved amino acid residues, and are responsible for  
25 proper polypeptide folding and for catalysis. The catalytic domain can be identified following a Smith-Waterman alignment of the protein sequence against the non-redundant protein database.

      The term "catalytic activity", as used herein, defines the rate at which a kinase catalytic domain phosphorylates a substrate. Catalytic activity can be measured, for example, by determining the amount of a substrate converted to a phosphorylated product as a function of  
30 time. Catalytic activity can be measured by methods of the invention by holding time constant

and determining the concentration of a phosphorylated substrate after a fixed period of time. Phosphorylation of a substrate occurs at the active site of a protein kinase. The active site is normally a cavity in which the substrate binds to the protein kinase and is phosphorylated.

The term "substrate" as used herein refers to a molecule phosphorylated by a kinase of the invention. Kinases phosphorylate substrates on serine/threonine or tyrosine amino acids. The molecule may be another protein or a polypeptide.

The term "C-terminal domain" refers to the region located between the catalytic domain or the last (located closest to the C-terminus) functional domain and the carboxy-terminal amino acid residue of the protein kinase. By "functional" domain is meant any region of the polypeptide that may play a regulatory or catalytic role as predicted from amino acid sequence homology to other proteins or by the presence of amino acid sequences that may give rise to specific structural conformations (*e.g.* N-terminal domain). The C-terminal domain can be identified by using a Smith-Waterman alignment of the protein sequence against the non-redundant protein database to define the C-terminal boundary of the catalytic domain or of any functional C-terminal extracatalytic domain. Depending on its length and amino acid composition, the C-terminal domain may or may not play a regulatory role in kinase function. An example of a protein kinase whose C-terminal domain may play a regulatory role is PAK3 which contains a heterotrimeric G<sub>b</sub> subunit-binding site near its C-terminus (Leeuw, T. *et al.* (1998) *Nature*, 391, 191-195). For the some of the kinases of the instant invention, the C-terminal domain may also comprise the catalytic domain (above).

The term "C-terminal tail" as used herein, refers to a C-terminal domain of a protein kinase, that by homology extends or protrudes past the C-terminal amino acid of its closest homolog. C-terminal tails can be identified by using a Smith-Waterman sequence alignment of the protein sequence against the non-redundant protein database, or by means of a multiple sequence alignment of homologous sequences using the DNASTar program Megalign. Depending on its length, a C-terminal tail may or may not play a regulatory role in kinase function.

The term "coiled-coil structure region" as used herein, refers to a polypeptide sequence that has a high probability of adopting a coiled-coil structure as predicted by computer algorithms such as COILS (Lupas, A. (1996) *Meth. Enzymology* 266:513-525). Coiled-coils are

formed by two or three amphipathic  $\alpha$ -helices in parallel. Coiled-coils can bind to coiled-coil domains of other polypeptides resulting in homo- or heterodimers (Lupas, A. (1991) *Science* 252:1162-1164). Coiled-coil-dependent oligomerization has been shown to be necessary for protein function including catalytic activity of serine/threonine kinases (Roe, J. *et al.* (1997) *J. Biol. Chem.* 272:5838-5845).

The term "proline-rich region" as used herein, refers to a region of a protein kinase whose proline content over a given amino acid length is higher than the average content of this amino acid found in proteins (*i.e.*, >10%). Proline-rich regions are easily discernable by visual inspection of amino acid sequences and quantitated by standard computer sequence analysis programs such as the DNASTar program EditSeq. Proline-rich regions have been demonstrated to participate in regulatory protein-protein interactions. Among these interactions, those that are most relevant to this invention involve the "PxxP" proline rich motif found in certain protein kinases (*i.e.*, human PAK1) and the SH3 domain of the adaptor molecule Nck (Galisteo, M.L. *et al.* (1996) *J. Biol. Chem.* 271:20997-21000). Other regulatory interactions involving "PxxP" proline-rich motifs include the WW domain (Sudol, M. (1996) *Prog. Biochys. Mol. Bio.* 65:113-132).

The term "spacer region" as used herein, refers to a region of the protein kinase located between predicted functional domains. The spacer region has no detectable homology to any amino acid sequence in the database, and can be identified by using a Smith-Waterman alignment of the protein sequence against the non-redundant protein database to define the C- and N-terminal boundaries of the flanking functional domains. Spacer regions may or may not play a fundamental role in protein kinase function. Precedence for the regulatory role of spacer regions in kinase function is provided by the role of the *src* kinase spacer in inter-domain interactions (Xu, W. *et al.* (1997) *Nature* 385:595-602).

The term "insert" as used herein refers to a portion of a protein kinase that is absent from a close homolog. Inserts may or may not be the product alternative splicing of exons. Inserts can be identified by using a Smith-Waterman sequence alignment of the protein sequence against the non-redundant protein database, or by means of a multiple sequence alignment of homologous sequences using the DNASTar program Megalign. Inserts may play a functional

role by presenting a new interface for protein-protein interactions, or by interfering with such interactions.

The term "signal transduction pathway" refers to the molecules that propagate an extracellular signal through the cell membrane to become an intracellular signal. This signal can then stimulate a cellular response. The polypeptide molecules involved in signal transduction processes are typically receptor and non-receptor protein tyrosine kinases, receptor and non-receptor protein phosphatases, polypeptides containing *SRC* homology 2 and 3 domains, phosphotyrosine binding proteins (*SRC* homology 2 (SH2) and phosphotyrosine binding (PTB and PH) domain containing proteins), proline-rich binding proteins (SH3 domain containing proteins), GTPases, phosphodiesterases, phospholipases, prolyl isomerases, proteases, Ca<sup>2+</sup>-binding proteins, cAMP binding proteins, guanyl cyclases, adenylyl cyclases, NO generating proteins, nucleotide exchange factors, and transcription factors.

In other preferred embodiments, the invention features isolated, enriched, or purified nucleic acid molecules encoding kinase polypeptides, further comprising a vector or promoter effective to initiate transcription in a host cell. The invention also features recombinant nucleic acid, preferably in a cell or an organism. The recombinant nucleic acid may contain a sequence selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57, or a functional derivative thereof and a vector or a promoter effective to initiate transcription in a host cell. The recombinant nucleic acid can alternatively contain a transcriptional initiation region functional in a cell, a sequence complementary to an RNA sequence encoding a kinase polypeptide and a transcriptional

termination region functional in a cell. Specific vectors and host cell combinations are discussed herein.

The term "vector" relates to a single or double-stranded circular nucleic acid molecule that can be transfected into cells and replicated within or independently of a cell genome. A  
5 circular double-stranded nucleic acid molecule can be cut and thereby linearized upon treatment with restriction enzymes. An assortment of nucleic acid vectors, restriction enzymes, and the knowledge of the nucleotide sequences cut by restriction enzymes are readily available to those skilled in the art. A nucleic acid molecule encoding a kinase can be inserted into a vector by cutting the vector with restriction enzymes and ligating the two pieces together.

10 The term "transfecting" defines a number of methods to insert a nucleic acid vector or other nucleic acid molecules into a cellular organism. These methods involve a variety of techniques, such as treating the cells with high concentrations of salt, an electric field, detergent, or DMSO to render the outer membrane or wall of the cells permeable to nucleic acid molecules of interest or use of various viral transduction strategies.

15 The term "promoter" as used herein, refers to nucleic acid sequence needed for gene sequence expression. Promoter regions vary from organism to organism, but are well known to persons skilled in the art for different organisms. For example, in prokaryotes, the promoter region contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such  
20 regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

In preferred embodiments, the isolated nucleic acid comprises, consists essentially of, or consists of a nucleic acid sequence selected from the group consisting of those set forth in SEQ  
25 ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID  
30 NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ



ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57, which encodes

5 an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID

10 NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ

15 ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, a functional derivative thereof, or at least 35, 40, 45, 50, 60, 75, 100, 200, or 300 contiguous amino acids selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ

20 ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID

25 NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. The nucleic acid may be isolated from a natural source by cDNA cloning or by subtractive hybridization. The natural source may be mammalian, preferably human, preferably

blood, semen or tissue, and the nucleic acid may be synthesized by the triester method or by using an automated DNA synthesizer.

The term "mammal" refers preferably to such organisms as mice, rats, rabbits, guinea pigs, sheep, and goats, more preferably to cats, dogs, monkeys, and apes, and most preferably to humans.

In yet other preferred embodiments, the nucleic acid is a conserved or unique region, for example those useful for: the design of hybridization probes to facilitate identification and cloning of additional polypeptides, the design of PCR probes to facilitate cloning of additional polypeptides, obtaining antibodies to polypeptide regions, and designing antisense oligonucleotides.

By "conserved nucleic acid regions", are meant regions present on two or more nucleic acids encoding a kinase polypeptide, to which a particular nucleic acid sequence can hybridize under lower stringency conditions. Examples of lower stringency conditions suitable for screening for nucleic acid encoding kinase polypeptides are provided in Wahl *et al. Meth. Enzym.* 152:399-407 (1987) and in Wahl *et al. Meth. Enzym.* 152:415-423 (1987), which are hereby incorporated by reference herein in its entirety, including any drawings, figures, or tables. Preferably, conserved regions differ by no more than 5 out of 20 nucleotides, even more preferably 2 out of 20 nucleotides or most preferably 1 out of 20 nucleotides.

By "unique nucleic acid region" is meant a sequence present in a nucleic acid coding for a kinase polypeptide that is not present in a sequence coding for any other naturally occurring polypeptide. Such regions preferably encode 32 (preferably 40, more preferably 45, most preferably 55) or more contiguous amino acids, for example, an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100,

SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. In particular, a unique nucleic acid region is preferably of mammalian origin.

- 5 Another aspect of the invention features a nucleic acid probe for the detection of nucleic acid encoding a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. in a sample. The nucleic acid probe contains a nucleotide base sequence that will hybridize to the sequence selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57, or a functional derivative thereof.

In preferred embodiments, the nucleic acid probe hybridizes to nucleic acid encoding at least 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids, wherein the nucleic acid sequence is selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57, or a functional derivative thereof.

Methods for using the probes include detecting the presence or amount of kinase RNA in a sample by contacting the sample with a nucleic acid probe under conditions such that hybridization occurs and detecting the presence or amount of the probe bound to kinase RNA. The nucleic acid duplex formed between the probe and a nucleic acid sequence coding for a kinase polypeptide may be used in the identification of the sequence of the nucleic acid detected (Nelson *et al.*, in *Nonisotopic DNA Probe Techniques*, Academic Press, San Diego, Kricka, ed., p. 275, 1992, hereby incorporated by reference herein in its entirety, including any drawings, figures, or tables). Kits for performing such methods may be constructed to include a container means having disposed therein a nucleic acid probe.

Methods for using the probes also include using these probes to find, for example, the full-length clone of each of the predicted kinases by techniques known to one skilled in the art. These clones will be useful for screening for small molecule compounds that inhibit the catalytic activity of the encoded kinase with potential utility in treating cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically disorders including cancers of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain,

sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, multiple sclerosis, and amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

In another aspect, the invention describes a recombinant cell or tissue comprising a nucleic acid molecule encoding a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. In such cells, the nucleic acid may be under the control of the genomic regulatory elements, or may be under the control of exogenous regulatory elements including an exogenous promoter. By "exogenous" it is meant a promoter that is not normally coupled *in vivo* transcriptionally to the coding sequence for the kinase polypeptides.

The polypeptide is preferably a fragment of the protein encoded by an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, 5 SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. By "fragment," is meant an amino acid sequence present in a kinase polypeptide. Preferably, such a sequence comprises at least 32, 45, 50, 60, 100, 200, or 300 contiguous amino 15 acids of a sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, 20 SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. 25

In another aspect, the invention features an isolated, enriched, or purified kinase polypeptide having the amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, 30

SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114.

10 By "isolated" in reference to a polypeptide is meant a polymer of 6 (preferably 12, more preferably 18, most preferably 25, 32, 40, or 50) or more amino acids conjugated to each other, including polypeptides that are isolated from a natural source or that are synthesized. In certain aspects longer polypeptides are preferred, such as those comprising 100, 200, 300, 400, 450, 500, 550, 600, 700, 800, 900 or more contiguous amino acids, including an amino acid sequence  
 15 selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID  
 20 NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114.  
 25

The isolated polypeptides of the present invention are unique in the sense that they are not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term

does not imply that the sequence is the only amino acid chain present, but that it is essentially free (about 90 - 95% pure at least) of non-amino acid-based material naturally associated with it.

By the use of the term "enriched" in reference to a polypeptide is meant that the specific amino acid sequence constitutes a significantly higher fraction (2- to 5-fold) of the total amino acid sequences present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other amino acid sequences present, or by a preferential increase in the amount of the specific amino acid sequence of interest, or by a combination of the two. However, it should be noted that enriched does not imply that there are no other amino acid sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term "significantly" here is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other amino acid sequences of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no amino acid sequence from other sources. The other source of amino acid sequences may, for example, comprise amino acid sequence encoded by a yeast or bacterial genome, or a cloning vector such as pUC19. The term is meant to cover only those situations in which man has intervened to increase the proportion of the desired amino acid sequence.

It is also advantageous for some purposes that an amino acid sequence be in purified form. The term "purified" in reference to a polypeptide does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment. Compared to the natural level this level should be at least 2-to 5-fold greater (*e.g.*, in terms of mg/mL). Purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. The substance is preferably free of contamination at a functionally significant level, for example 90%, 95%, or 99% pure.

In preferred embodiments, the kinase polypeptide is a fragment of the protein encoded by an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ



ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. Preferably, the kinase polypeptide contains at least 32, 45, 50, 60, 100, 200, or 300 contiguous amino acids of a sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, or a functional derivative thereof.

In preferred embodiments, the kinase polypeptide comprises an amino acid sequence having (a) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97,

SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114; and (b) an amino acid sequence selected from the group

5 consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID

10 NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID

15 NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, except that it lacks one or more of the domains selected from the group consisting of a C-terminal catalytic domain, an N-terminal domain, a catalytic domain, a C-terminal domain, a coiled-coil structure region, a proline-rich region, a spacer region, and a C-terminal tail.

The polypeptide can be isolated from a natural source by methods well-known in the art.

20 The natural source may be mammalian, preferably human, preferably blood, semen or tissue, and the polypeptide may be synthesized using an automated polypeptide synthesizer.

In some embodiments the invention includes a recombinant kinase polypeptide having (a) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID

25

30

NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. By "recombinant kinase polypeptide" is meant a polypeptide produced by recombinant

- 5 DNA techniques such that it is distinct from a naturally occurring polypeptide either in its location (*e.g.*, present in a different cell or tissue than found in nature), purity or structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

- The polypeptides to be expressed in host cells may also be fusion proteins which include  
10 regions from heterologous proteins. Such regions may be included to allow, *e.g.*, secretion, improved stability, or facilitated purification of the polypeptide. For example, a sequence encoding an appropriate signal peptide can be incorporated into expression vectors. A DNA sequence for a signal peptide (secretory leader) may be fused in-frame to the polynucleotide sequence so that the polypeptide is translated as a fusion protein comprising the signal peptide.  
15 A signal peptide that is functional in the intended host cell promotes extracellular secretion of the polypeptide. Preferably, the signal sequence will be cleaved from the polypeptide upon secretion of the polypeptide from the cell. Thus, preferred fusion proteins can be produced in which the N-terminus of a kinase polypeptide is fused to a carrier peptide.

- In one embodiment, the polypeptide comprises a fusion protein which includes a  
20 heterologous region used to facilitate purification of the polypeptide. Many of the available peptides used for such a function allow selective binding of the fusion protein to a binding partner. A preferred binding partner includes one or more of the IgG binding domains of protein A are easily purified to homogeneity by affinity chromatography on, for example, IgG-coupled Sepharose. Alternatively, many vectors have the advantage of carrying a stretch of histidine  
25 residues that can be expressed at the N-terminal or C-terminal end of the target protein, and thus the protein of interest can be recovered by metal chelation chromatography. A nucleotide sequence encoding a recognition site for a proteolytic enzyme such as enterokinase, factor X procollagenase or thrombin may immediately precede the sequence for a kinase polypeptide to permit cleavage of the fusion protein to obtain the mature kinase polypeptide. Additional  
30 examples of fusion-protein binding partners include, but are not limited to, the yeast I-factor, the

honeybee melatin leader in sf9 insect cells, 6-His tag, thioredoxin tag, hemagglutinin tag, GST tag, and OmpA signal sequence tag. As will be understood by one of skill in the art, the binding partner which recognizes and binds to the peptide may be any ion, molecule or compound including metal ions (*e.g.*, metal affinity columns), antibodies, or fragments thereof, and any  
5 protein or peptide which binds the peptide, such as the FLAG tag.

In another aspect, the invention features an antibody (*e.g.*, a monoclonal or polyclonal antibody) having specific binding affinity to a kinase polypeptide or a kinase polypeptide domain or fragment where the polypeptide is selected from the group having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID  
10 NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88,  
15 SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. By  
20 “specific binding affinity” is meant that the antibody binds to the target kinase polypeptide with greater affinity than it binds to other polypeptides under specified conditions. Antibodies or antibody fragments are polypeptides that contain regions that can bind other polypeptides. The term “specific binding affinity” describes an antibody that binds to a kinase polypeptide with greater affinity than it binds to other polypeptides under specified conditions.. Antibodies can be  
25 used to identify an endogenous source of kinase polypeptides, to monitor cell cycle regulation, and for immuno-localization of kinase polypeptides within the cell.

The term “polyclonal” refers to antibodies that are heterogenous populations of antibody molecules derived from the sera of animals immunized with an antigen or an antigenic functional derivative thereof. For the production of polyclonal antibodies, various host animals may be

immunized by injection with the antigen. Various adjuvants may be used to increase the immunological response, depending on the host species.

“Monoclonal antibodies” are substantially homogenous populations of antibodies to a particular antigen. They may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. Monoclonal antibodies may be obtained  
5 by methods known to those skilled in the art (Kohler *et al.*, *Nature* 256:495-497, 1975, and U.S. Patent No. 4,376,110, both of which are hereby incorporated by reference herein in their entirety including any figures, tables, or drawings).

The term “antibody fragment” refers to a portion of an antibody, often the hypervariable  
10 region and portions of the surrounding heavy and light chains, that displays specific binding affinity for a particular molecule. A hypervariable region is a portion of an antibody that physically binds to the polypeptide target.

Antibodies or antibody fragments having specific binding affinity to a kinase polypeptide of the invention may be used in methods for detecting the presence and/or amount of kinase  
15 polypeptide in a sample by probing the sample with the antibody under conditions suitable for kinase-antibody immunocomplex formation and detecting the presence and/or amount of the antibody conjugated to the kinase polypeptide. Diagnostic kits for performing such methods may be constructed to include antibodies or antibody fragments specific for the kinase as well as a conjugate of a binding partner of the antibodies or the antibodies themselves.

20 An antibody or antibody fragment with specific binding affinity to a kinase polypeptide of the invention can be isolated, enriched, or purified from a prokaryotic or eukaryotic organism. Routine methods known to those skilled in the art enable production of antibodies or antibody fragments, in both prokaryotic and eukaryotic organisms. Purification, enrichment, and isolation of antibodies, which are polypeptide molecules, are described above.

25 Antibodies having specific binding affinity to a kinase polypeptide of the invention may be used in methods for detecting the presence and/or amount of kinase polypeptide in a sample by contacting the sample with the antibody under conditions such that an immunocomplex forms and detecting the presence and/or amount of the antibody conjugated to the kinase polypeptide. Diagnostic kits for performing such methods may be constructed to include a first container  
30 containing the antibody and a second container having a conjugate of a binding partner of the

antibody and a label, such as, for example, a radioisotope. The diagnostic kit may also include notification of an FDA approved use and instructions therefor.

In another aspect, the invention features a hybridoma which produces an antibody having specific binding affinity to a kinase polypeptide or a kinase polypeptide domain, where the polypeptide is selected from the group having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. By "hybridoma" is meant an immortalized cell line that is capable of secreting an antibody, for example an antibody to a kinase of the invention. In preferred embodiments, the antibody to the kinase comprises a sequence of amino acids that is able to specifically bind a kinase polypeptide of the invention.

In another aspect, the present invention is also directed to kits comprising antibodies that bind to a polypeptide encoded by any of the nucleic acid molecules described above, and a negative control antibody.

The term "negative control antibody" refers to an antibody derived from similar source as the antibody having specific binding affinity, but where it displays no binding affinity to a polypeptide of the invention.

In another aspect, the invention features a kinase polypeptide binding agent able to bind to a kinase polypeptide selected from the group having (a) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID

NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. The binding agent is preferably a purified antibody that recognizes an epitope present on a kinase polypeptide of the invention. Other binding agents include molecules that bind to kinase polypeptides and analogous molecules that bind to a kinase polypeptide. Such binding agents may be identified by using assays that measure kinase binding partner activity, such as those that measure PDGFR activity.

The invention also features a method for screening for human cells containing a kinase polypeptide of the invention or an equivalent sequence. The method involves identifying the novel polypeptide in human cells using techniques that are routine and standard in the art, such as those described herein for identifying the kinases of the invention (*e.g.*, cloning, Southern or Northern blot analysis, *in situ* hybridization, PCR amplification, etc.).

In another aspect, the invention features methods for identifying a substance that modulates kinase activity comprising the steps of: (a) contacting a kinase polypeptide selected from the group having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID

NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114 with a test substance; (b) measuring the activity of said polypeptide; and (c) determining whether said substance modulates the activity of said polypeptide. The skilled artisan will appreciate that the kinase polypeptides of the invention, including, for example, a portion of a full-length sequence such as a catalytic domain or a portion thereof, are useful for the identification of a substance which modulates kinase activity. Those kinase polypeptides having a functional activity (*e.g.*, catalytic activity as defined herein) are useful for identifying a substance that modulates kinase activity.

The term “modulates” refers to the ability of a compound to alter the function of a kinase of the invention. A modulator preferably activates or inhibits the activity of a kinase of the invention depending on the concentration of the compound exposed to the kinase.

The term “modulates” also refers to altering the function of kinases of the invention by increasing or decreasing the probability that a complex forms between the kinase and a natural binding partner. A modulator preferably increases the probability that such a complex forms between the kinase and the natural binding partner, more preferably increases or decreases the probability that a complex forms between the kinase and the natural binding partner depending on the concentration of the compound exposed to the kinase, and most preferably decreases the probability that a complex forms between the kinase and the natural binding partner.

The term “activates” refers to increasing the cellular activity of the kinase. The term inhibit refers to decreasing the cellular activity of the kinase. Kinase activity is preferably the interaction with a natural binding partner.

The term “complex” refers to an assembly of at least two molecules bound to one another. Signal transduction complexes often contain at least two protein molecules bound to one another. For instance, a protein tyrosine receptor protein kinase, GRB2, SOS, RAF, and RAS assemble to form a signal transduction complex in response to a mitogenic ligand.

The term “natural binding partner” refers to polypeptides, lipids, small molecules, or nucleic acids that bind to kinases in cells. A change in the interaction between a kinase and a natural binding partner can manifest itself as an increased or decreased probability that the interaction forms, or an increased or decreased concentration of kinase/natural binding partner complex.



The term “contacting” as used herein refers to mixing a solution comprising the test compound with a liquid medium bathing the cells of the methods. The solution comprising the compound may also comprise another component, such as dimethyl sulfoxide (DMSO), which facilitates the uptake of the test compound or compounds into the cells of the methods. The  
5 solution comprising the test compound may be added to the medium bathing the cells by utilizing a delivery apparatus, such as a pipette-based device or syringe-based device.

In another aspect, the invention features methods for identifying a substance that modulates kinase activity in a cell comprising the steps of: (a) expressing a kinase polypeptide in a cell, wherein said polypeptide is selected from the group having an amino acid sequence  
10 selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ  
15 ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID  
20 NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114; (b) adding a test substance to said cell; and (c) monitoring a change in cell phenotype or the interaction between said polypeptide and a natural binding partner. The skilled artisan will appreciate that the kinase polypeptides of the invention, including, for example, a portion of a full-length sequence such as a catalytic domain or a portion thereof, are useful for the identification of a  
25 substance which modulates kinase activity. Those kinase polypeptides having a functional activity (*e.g.*, catalytic activity as defined herein) are useful for identifying a substance that modulates kinase activity.

The term “expressing” as used herein refers to the production of kinases of the invention from a nucleic acid vector containing kinase genes within a cell. The nucleic acid vector is  
30 transfected into cells using well known techniques in the art as described herein.

Another aspect of the instant invention is directed to methods of identifying compounds that bind to kinase polypeptides of the present invention, comprising contacting the kinase polypeptides with a compound, and determining whether the compound binds the kinase polypeptides. Binding can be determined by binding assays which are well known to the skilled artisan, including, but not limited to, gel-shift assays, Western blots, radiolabeled competition  
5 assay, phage-based expression cloning, co-fractionation by chromatography, co-precipitation, cross linking, interaction trap/two-hybrid analysis, southwestern analysis, ELISA, and the like, which are described in, for example, *Current Protocols in Molecular Biology*, 1999, John Wiley & Sons, NY, which is incorporated herein by reference in its entirety. The compounds to be  
10 screened include, but are not limited to, compounds of extracellular, intracellular, biological or chemical origin.

The methods of the invention also embrace compounds that are attached to a label, such as a radiolabel (*e.g.*,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^3\text{H}$ ), a fluorescence label, a chemiluminescent label, an enzymic label and an immunogenic label. The kinase polypeptides employed in such a test may  
15 either be free in solution, attached to a solid support, borne on a cell surface, located intracellularly or associated with a portion of a cell. One skilled in the art can, for example, measure the formation of complexes between a kinase polypeptide and the compound being tested. Alternatively, one skilled in the art can examine the diminution in complex formation between a kinase polypeptide and its substrate caused by the compound being tested.

20 Other assays can be used to examine enzymatic activity including, but not limited to, photometric, radiometric, HPLC, electrochemical, and the like, which are described in, for example, *Enzyme Assays: A Practical Approach*, eds. R. Eisenthal and M. J. Danson, 1992, Oxford University Press, which is incorporated herein by reference in its entirety.

Another aspect of the present invention is directed to methods of identifying compounds  
25 which modulate (*i.e.*, increase or decrease) activity of a kinase polypeptide comprising contacting the kinase polypeptide with a compound, and determining whether the compound modifies activity of the kinase polypeptide. As described herein, the kinase polypeptides of the invention include a portion of a full-length sequence, such as a catalytic domain, as defined herein. In some instances, the kinase polypeptides of the invention comprise less than the entire  
30 catalytic domain, yet exhibit kinase or kinase-like activity. These compounds are also referred to

as "modulators of protein kinases." The activity in the presence of the test compound is measured to the activity in the absence of the test compound. Where the activity of a sample containing the test compound is higher than the activity in a sample lacking the test compound, the compound will have increased the activity. Similarly, where the activity of a sample  
5 containing the test compound is lower than the activity in the sample lacking the test compound, the compound will have inhibited the activity.

The present invention is particularly useful for screening compounds by using a kinase polypeptide in any of a variety of drug screening techniques. The compounds to be screened include, but are not limited to, extracellular, intracellular, biological or chemical origin. The  
10 kinase polypeptide employed in such a test may be in any form, preferably, free in solution, attached to a solid support, borne on a cell surface or located intracellularly. One skilled in the art can, for example, measure the formation of complexes between a kinase polypeptide and the compound being tested. Alternatively, one skilled in the art can examine the diminution in complex formation between a kinase polypeptide and its substrate caused by the compound  
15 being tested.

The activity of kinase polypeptides of the invention can be determined by, for example, examining the ability to bind or be activated by chemically synthesised peptide ligands. Alternatively, the activity of the kinase polypeptides can be assayed by examining their ability to bind metal ions such as calcium, hormones, chemokines, neuropeptides, neurotransmitters,  
20 nucleotides, lipids, odorants, and photons. Thus, modulators of the kinase polypeptide's activity may alter a kinase function, such as a binding property of a kinase or an activity such as signal transduction or membrane localization.

In various embodiments of the method, the assay may take the form of a yeast growth assay, an Aequorin assay, a Luciferase assay, a mitogenesis assay, a MAP Kinase activity assay,  
25 as well as other binding or function-based assays of kinase activity that are generally known in the art. In several of these embodiments, the invention includes any of the receptor and non-receptor protein tyrosine kinases, receptor and non-receptor protein phosphatases, polypeptides containing *SRC* homology 2 and 3 domains, phosphotyrosine binding proteins (*SRC* homology 2 (SH2) and phosphotyrosine binding (PTB and PH) domain containing proteins), proline-rich  
30 binding proteins (SH3 domain containing proteins), GTPases, phosphodiesterases,

phospholipases, prolyl isomerases, proteases, Ca<sup>2+</sup> binding proteins, cAMP binding proteins, guanyl cyclases, adenylyl cyclases, NO generating proteins, nucleotide exchange factors, and transcription factors. Biological activities of kinases according to the invention include, but are not limited to, the binding of a natural or a synthetic ligand, as well as any one of the functional activities of kinases known in the art. Non-limiting examples of kinase activities include  
5 transmembrane signaling of various forms, which may involve kinase binding interactions and/or the exertion of an influence over signal transduction.

The modulators of the invention exhibit a variety of chemical structures, which can be generally grouped into mimetics of natural kinase ligands, and peptide and non-peptide allosteric effectors of kinases. The invention does not restrict the sources for suitable modulators, which  
10 may be obtained from natural sources such as plant, animal or mineral extracts, or non-natural sources such as small molecule libraries, including the products of combinatorial chemical approaches to library construction, and peptide libraries.

The use of cDNAs encoding kinases in drug discovery programs is well-known; assays  
15 capable of testing thousands of unknown compounds per day in high-throughput screens (HTSs) are thoroughly documented. The literature is replete with examples of the use of radiolabelled ligands in HTS binding assays for drug discovery (see Williams, *Medicinal Research Reviews*, 1991, 11, 147-184.; Sweetnam, *et al.*, *J. Natural Products*, 1993, 56, 441-455 for review). Recombinant receptors are preferred for binding assay HTS because they allow for better  
20 specificity (higher relative purity), provide the ability to generate large amounts of receptor material, and can be used in a broad variety of formats (see Hodgson, *Bio/Technology*, 1992, 10, 973-980; each of which is incorporated herein by reference in its entirety).

A variety of heterologous systems is available for functional expression of recombinant receptors that are well known to those skilled in the art. Such systems include bacteria  
25 (Strosberg, *et al.*, *Trends in Pharmacological Sciences*, 1992, 13, 95-98), yeast (Pausch, *Trends in Biotechnology*, 1997, 15, 487-494), several kinds of insect cells (Vanden Broeck, *Int. Rev. Cytology*, 1996, 164, 189-268), amphibian cells (Jayawickreme *et al.*, *Current Opinion in Biotechnology*, 1997, 8, 629-634) and several mammalian cell lines (CHO, HEK293, COS, etc.; see Gerhardt, *et al.*, *Eur. J. Pharmacology*, 1997, 334, 1-23). These examples do not preclude

the use of other possible cell expression systems, including cell lines obtained from nematodes (PCT application WO 98/37177).

An expressed kinase can be used for HTS binding assays in conjunction with its defined ligand, in this case the corresponding peptide that activates it. The identified peptide is labeled  
5 with a suitable radioisotope, including, but not limited to,  $^{125}\text{I}$ ,  $^3\text{H}$ ,  $^{35}\text{S}$  or  $^{32}\text{P}$ , by methods that are well known to those skilled in the art. Alternatively, the peptides may be labeled by well-known methods with a suitable fluorescent derivative (Baindur, *et al.*, *Drug Dev. Res.*, 1994, 33, 373-398; Rogers, *Drug Discovery Today*, 1997, 2, 156-160). Radioactive ligand specifically bound to the receptor in membrane preparations made from the cell line expressing the recombinant  
10 protein can be detected in HTS assays in one of several standard ways, including filtration of the receptor-ligand complex to separate bound ligand from unbound ligand (Williams, *Med. Res. Rev.*, 1991, 11, 147-184.; Sweetnam, *et al.*, *J. Natural Products*, 1993, 56, 441-455). Alternative methods include a scintillation proximity assay (SPA) or a FlashPlate format in which such separation is unnecessary (Nakayama, *Cur. Opinion Drug Disc. Dev.*, 1998, 1, 85-91 Bossé, *et al.*, *J. Biomolecular Screening*, 1998, 3, 285-292.). Binding of fluorescent ligands can be  
15 detected in various ways, including fluorescence energy transfer (FRET), direct spectrophotofluorometric analysis of bound ligand, or fluorescence polarization (Rogers, *Drug Discovery Today*, 1997, 2, 156-160; Hill, *Cur. Opinion Drug Disc. Dev.*, 1998, 1, 92-97).

The kinases and natural binding partners required for functional expression of  
20 heterologous kinase polypeptides can be native constituents of the host cell or can be introduced through well-known recombinant technology. The kinase polypeptides can be intact or chimeric. The kinase activation results in the stimulation or inhibition of other native proteins, events that can be linked to a measurable response.

Examples of such biological responses include, but are not limited to, the following: the  
25 ability to survive in the absence of a limiting nutrient in specifically engineered yeast cells (Pausch, *Trends in Biotechnology*, 1997, 15, 487-494); changes in intracellular  $\text{Ca}^{2+}$  concentration as measured by fluorescent dyes (Murphy, *et al.*, *Cur. Opinion Drug Disc. Dev.*, 1998, 1, 192-199). Fluorescence changes can also be used to monitor ligand-induced changes in membrane potential or intracellular pH; an automated system suitable for HTS has been  
30 described for these purposes (Schroeder, *et al.*, *J. Biomolecular Screening*, 1996, 1, 75-80).

Assays are also available for the measurement of common second but these are not generally preferred for HTS.

The invention contemplates a multitude of assays to screen and identify inhibitors of ligand binding to kinase polypeptides. In one example, the kinase polypeptide is immobilized and interaction with a binding partner is assessed in the presence and absence of a candidate modulator such as an inhibitor compound. In another example, interaction between the kinase polypeptide and its binding partner is assessed in a solution assay, both in the presence and absence of a candidate inhibitor compound. In either assay, an inhibitor is identified as a compound that decreases binding between the kinase polypeptide and its natural binding partner. Another contemplated assay involves a variation of the di-hybrid assay wherein an inhibitor of protein/protein interactions is identified by detection of a positive signal in a transformed or transfected host cell, as described in PCT publication number WO 95/20652, published August 3, 1995 and is included by reference herein including any figures, tables, or drawings.

Candidate modulators contemplated by the invention include compounds selected from libraries of either potential activators or potential inhibitors. There are a number of different libraries used for the identification of small molecule modulators, including: (1) chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules. Chemical libraries consist of random chemical structures, some of which are analogs of known compounds or analogs of compounds that have been identified as "hits" or "leads" in other drug discovery screens, while others are derived from natural products, and still others arise from non-directed synthetic organic chemistry. Natural product libraries are collections of microorganisms, animals, plants, or marine organisms which are used to create mixtures for screening by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of plants or marine organisms. Natural product libraries include polyketides, non-ribosomal peptides, and variants (non-naturally occurring) thereof. For a review, see *Science* 282:63-68 (1998). Combinatorial libraries are composed of large numbers of peptides, oligonucleotides, or organic compounds as a mixture. These libraries are relatively easy to prepare by traditional automated synthesis methods, PCR, cloning, or proprietary synthetic methods. Of particular interest are non-peptide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel

synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to modulate activity.

Still other candidate inhibitors contemplated by the invention can be designed and include soluble forms of binding partners, as well as such binding partners as chimeric, or fusion, proteins. A "binding partner" as used herein broadly encompasses both natural binding partners as described above as well as chimeric polypeptides, peptide modulators other than natural ligands, antibodies, antibody fragments, and modified compounds comprising antibody domains that are immunospecific for the expression product of the identified kinase gene.

Other assays may be used to identify specific peptide ligands of a kinase polypeptide, including assays that identify ligands of the target protein through measuring direct binding of test ligands to the target protein, as well as assays that identify ligands of target proteins through affinity ultrafiltration with ion spray mass spectroscopy/HPLC methods or other physical and analytical methods. Alternatively, such binding interactions are evaluated indirectly using the yeast two-hybrid system described in Fields *et al.*, *Nature*, 340:245-246 (1989), and Fields *et al.*, *Trends in Genetics*, 10:286-292 (1994), both of which are incorporated herein by reference. The two-hybrid system is a genetic assay for detecting interactions between two proteins or polypeptides. It can be used to identify proteins that bind to a known protein of interest, or to delineate domains or residues critical for an interaction. Variations on this methodology have been developed to clone genes that encode DNA binding proteins, to identify peptides that bind to a protein, and to screen for drugs. The two-hybrid system exploits the ability of a pair of interacting proteins to bring a transcription activation domain into close proximity with a DNA binding domain that binds to an upstream activation sequence (UAS) of a reporter gene, and is generally performed in yeast. The assay requires the construction of two hybrid genes encoding (1) a DNA-binding domain that is fused to a first protein and (2) an activation domain fused to a second protein. The DNA-binding domain targets the first hybrid protein to the UAS of the reporter gene; however, because most proteins lack an activation domain, this DNA-binding hybrid protein does not activate transcription of the reporter gene. The second hybrid protein,

which contains the activation domain, cannot by itself activate expression of the reporter gene because it does not bind the UAS. However, when both hybrid proteins are present, the noncovalent interaction of the first and second proteins tethers the activation domain to the UAS, activating transcription of the reporter gene. For example, when the first protein is a kinase gene product, or fragment thereof, that is known to interact with another protein or nucleic acid, this assay can be used to detect agents that interfere with the binding interaction. Expression of the reporter gene is monitored as different test agents are added to the system. The presence of an inhibitory agent results in lack of a reporter signal.

When the function of the kinase polypeptide gene product is unknown and no ligands are known to bind the gene product, the yeast two-hybrid assay can also be used to identify proteins that bind to the gene product. In an assay to identify proteins that bind to a kinase polypeptide, or fragment thereof, a fusion polynucleotide encoding both a kinase polypeptide (or fragment) and a UAS binding domain (*i.e.*, a first protein) may be used. In addition, a large number of hybrid genes each encoding a different second protein fused to an activation domain are produced and screened in the assay. Typically, the second protein is encoded by one or more members of a total cDNA or genomic DNA fusion library, with each second protein coding region being fused to the activation domain. This system is applicable to a wide variety of proteins, and it is not even necessary to know the identity or function of the second binding protein. The system is highly sensitive and can detect interactions not revealed by other methods; even transient interactions may trigger transcription to produce a stable mRNA that can be repeatedly translated to yield the reporter protein.

Other assays may be used to search for agents that bind to the target protein. One such screening method to identify direct binding of test ligands to a target protein is described in U.S. Patent No. 5,585,277, incorporated herein by reference. This method relies on the principle that proteins generally exist as a mixture of folded and unfolded states, and continually alternate between the two states. When a test ligand binds to the folded form of a target protein (*i.e.*, when the test ligand is a ligand of the target protein), the target protein molecule bound by the ligand remains in its folded state. Thus, the folded target protein is present to a greater extent in the presence of a test ligand which binds the target protein, than in the absence of a ligand. Binding of the ligand to the target protein can be determined by any method which distinguishes



between the folded and unfolded states of the target protein. The function of the target protein need not be known in order for this assay to be performed. Virtually any agent can be assessed by this method as a test ligand, including, but not limited to, metals, polypeptides, proteins, lipids, polysaccharides, polynucleotides and small organic molecules.

5           Another method for identifying ligands of a target protein is described in Wieboldt *et al.*, *Anal. Chem.*, 69:1683-1691 (1997), incorporated herein by reference. This technique screens combinatorial libraries of 20-30 agents at a time in solution phase for binding to the target protein. Agents that bind to the target protein are separated from other library components by simple membrane washing. The specifically selected molecules that are retained on the filter are  
10 subsequently liberated from the target protein and analyzed by HPLC and pneumatically assisted electrospray (ion spray) ionization mass spectroscopy. This procedure selects library components with the greatest affinity for the target protein, and is particularly useful for small molecule libraries.

          In preferred embodiments of the invention, methods of screening for compounds which  
15 modulate kinase activity comprise contacting test compounds with kinase polypeptides and assaying for the presence of a complex between the compound and the kinase polypeptide. In such assays, the ligand is typically labelled. After suitable incubation, free ligand is separated from that present in bound form, and the amount of free or uncomplexed label is a measure of the ability of the particular compound to bind to the kinase polypeptide.

20           In another embodiment of the invention, high throughput screening for compounds having suitable binding affinity to kinase polypeptides is employed. Briefly, large numbers of different small peptide test compounds are synthesised on a solid substrate. The peptide test compounds are contacted with the kinase polypeptide and washed. Bound kinase polypeptide is then detected by methods well known in the art. Purified polypeptides of the invention can also  
25 be coated directly onto plates for use in the aforementioned drug screening techniques. In addition, non-neutralizing antibodies can be used to capture the protein and immobilize it on the solid support.

          Other embodiments of the invention comprise using competitive screening assays in which neutralizing antibodies capable of binding a polypeptide of the invention specifically  
30 compete with a test compound for binding to the polypeptide. In this manner, the antibodies can

be used to detect the presence of any peptide that shares one or more antigenic determinants with a kinase polypeptide. Radiolabeled competitive binding studies are described in A.H. Lin *et al.* *Antimicrobial Agents and Chemotherapy*, 1997, vol. 41, no. 10. pp. 2127-2131, the disclosure of which is incorporated herein by reference in its entirety.

5 In another aspect, the invention provides methods for treating a disease by administering to a patient in need of such treatment a substance that modulates the activity of a kinase polypeptide selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, 10 SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, as well as the full-length polypeptide thereof, or a portion of any of these sequences that retains functional activity, as described herein. Preferably the disease is selected from the group 20 consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention 25 disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other 30 viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes

and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory  
5 bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

In preferred embodiments, the invention provides methods for treating or preventing a disease or disorder by administering to a patient in need of such treatment a substance that modulates the activity of a kinase polypeptide having an amino acid sequence selected from the  
10 group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83,  
15 SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, as well as the full-length polypeptide thereof, or a portion of any of these sequences that retains functional activity, as described herein. Preferably, the disease is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues,  
25 blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and  
30 dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases

including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

10       The invention also features methods of treating or preventing a disease or disorder by administering to a patient in need of such treatment a substance that modulates the activity of a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, as well as the full-length polypeptide thereof, or a portion of any of these sequences that retains functional activity, as described herein.

25       Preferably the disease is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and

hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by

5 HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis,

10 chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

The invention also features methods of treating or preventing a disease or disorder by administering to a patient in need of such treatment a substance that modulates the activity of a

15 kinase polypeptide having an amino acid sequence selected from the group consisting those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ

20 ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID

25 NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, as well as the full-length polypeptide thereof, or a portion of any of these sequences that retains functional activity, as described herein.

Preferably the disease is selected from the group consisting of immune-related diseases and disorders, cardiovascular disease, and cancer. More preferably these diseases include cancer of

30 tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate,

cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental  
5 retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including  
10 reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection. Most  
15 preferably, the immune-related diseases and disorders are selected from the group consisting of rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplantation.

Substances useful for treatment of kinase-related disorders or diseases preferably show  
20 positive results in one or more *in vitro* assays for an activity corresponding to treatment of the disease or disorder in question (Examples of such assays are provided in the references in section VI, below; and in Example 7, herein). Examples of substances that can be screened for favorable activity are provided and referenced in section VI, below. The substances that modulate the activity of the kinases preferably include, but are not limited to, antisense oligonucleotides and  
25 inhibitors of protein kinases, as determined by methods and screens referenced in section VI and Example 7, below.

The term "preventing" refers to decreasing the probability that an organism contracts or develops an abnormal condition.

The term "treating" refers to having a therapeutic effect and at least partially alleviating  
30 or abrogating an abnormal condition in the organism.

The term "therapeutic effect" refers to the inhibition or activation factors causing or contributing to the abnormal condition. A therapeutic effect relieves to some extent one or more of the symptoms of the abnormal condition. In reference to the treatment of abnormal conditions, a therapeutic effect can refer to one or more of the following: (a) an increase in the proliferation, growth, and/or differentiation of cells; (b) inhibition (*i.e.*, slowing or stopping) of cell death; (c) inhibition of degeneration; (d) relieving to some extent one or more of the symptoms associated with the abnormal condition; and (e) enhancing the function of the affected population of cells. Compounds demonstrating efficacy against abnormal conditions can be identified as described herein.

10       The term "abnormal condition" refers to a function in the cells or tissues of an organism that deviates from their normal functions in that organism. An abnormal condition can relate to cell proliferation, cell differentiation, or cell survival.

Abnormal cell proliferative conditions include cancers such as fibrotic and mesangial disorders, abnormal angiogenesis and vasculogenesis, wound healing, psoriasis, diabetes mellitus, and inflammation.

Abnormal differentiation conditions include, but are not limited to neurodegenerative disorders, slow wound healing rates, and slow tissue grafting healing rates.

Abnormal cell survival conditions relate to conditions in which programmed cell death (apoptosis) pathways are activated or abrogated. A number of protein kinases are associated with the apoptosis pathways. Aberrations in the function of any one of the protein kinases could lead to cell immortality or premature cell death.

The term "aberration", in conjunction with the function of a kinase in a signal transduction process, refers to a kinase that is over- or under-expressed in an organism, mutated such that its catalytic activity is lower or higher than wild-type protein kinase activity, mutated such that it can no longer interact with a natural binding partner, is no longer modified by another protein kinase or protein phosphatase, or no longer interacts with a natural binding partner.

The term "administering" relates to a method of incorporating a compound into cells or tissues of an organism. The abnormal condition can be prevented or treated when the cells or tissues of the organism exist within the organism or outside of the organism. Cells existing

outside the organism can be maintained or grown in cell culture dishes. For cells harbored within the organism, many techniques exist in the art to administer compounds, including (but not limited to) oral, parenteral, dermal, injection, and aerosol applications. For cells outside of the organism, multiple techniques exist in the art to administer the compounds, including (but not limited to) cell microinjection techniques, transformation techniques, and carrier techniques.

The abnormal condition can also be prevented or treated by administering a compound to a group of cells having an aberration in a signal transduction pathway to an organism. The effect of administering a compound on organism function can then be monitored. The organism is preferably a mouse, rat, rabbit, guinea pig, or goat, more preferably a monkey or ape, and most preferably a human.

In another aspect, the invention features methods for detection of a kinase polypeptide in a sample as a diagnostic tool for diseases or disorders, wherein the method comprises the steps of: (a) contacting the sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid target region of a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, said probe comprising the nucleic acid sequence encoding the polypeptide, fragments thereof, and the complements of the sequences and fragments; and (b) detecting the presence or amount of the probe:target region hybrid as an indication of the disease.

In preferred embodiments of the invention, the disease or disorder is selected from the group consisting of rheumatoid arthritis, arteriosclerosis, autoimmune disorders, organ



transplantation, myocardial infarction, cardiomyopathies, stroke, renal failure, oxidative stress-related neurodegenerative disorders, and cancer.

The kinase "target region" is the nucleotide base sequence selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57, or the corresponding full-length sequences, a functional derivative thereof, or a fragment thereof, to which the nucleic acid probe will specifically hybridize. Specific hybridization indicates that in the presence of other nucleic acids the probe only hybridizes detectably with the kinase of the invention's target region. Putative target regions can be identified by methods well known in the art consisting of alignment and comparison of the most closely related sequences in the database.

In preferred embodiments the nucleic acid probe hybridizes to a kinase target region encoding at least 6, 12, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids of a sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID

NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, or the corresponding full-length amino acid sequence, a portion of any of these sequences that retains functional activity, as described herein, or a functional derivative thereof.

- 5 Hybridization conditions should be such that hybridization occurs only with the kinase genes in the presence of other nucleic acid molecules. Under stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having more than 1 or 2 mismatches out of 20 contiguous nucleotides. Such conditions are defined *supra*.

- 10 The diseases for which detection of kinase genes in a sample could be diagnostic include diseases in which kinase nucleic acid (DNA and/or RNA) is amplified in comparison to normal cells. By "amplification" is meant increased numbers of kinase DNA or RNA in a cell compared with normal cells. In normal cells, kinases are typically found as single copy genes. In selected diseases, the chromosomal location of the kinase genes may be amplified, resulting in multiple  
15 copies of the gene, or amplification. Gene amplification can lead to amplification of kinase RNA, or kinase RNA can be amplified in the absence of kinase DNA amplification.

- "Amplification" as it refers to RNA can be the detectable presence of kinase RNA in cells, since in some normal cells there is no basal expression of kinase RNA. In other normal cells, a basal level of expression of kinase exists, therefore in these cases amplification is the  
20 detection of at least 1-2-fold, and preferably more, kinase RNA, compared to the basal level.

- The diseases that could be diagnosed by detection of kinase nucleic acid in a sample preferably include cancers. The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the  
25 detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample that is compatible with the method utilized.

- The invention also features a method for detection of a kinase polypeptide in a sample as a diagnostic tool for a disease or disorder, wherein the method comprises: (a) comparing a  
30 nucleic acid target region encoding the kinase polypeptide in a sample, where the kinase

polypeptide has an amino acid sequence selected from the group consisting those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, or one or more fragments thereof, with a control nucleic acid target region encoding the kinase polypeptide, or one or more fragments thereof; and (b) detecting differences in sequence or amount between the target region and the control target region, as an indication of the disease or disorder. Preferably the disease is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis,

chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

5       The term “comparing” as used herein refers to identifying discrepancies between the nucleic acid target region isolated from a sample, and the control nucleic acid target region. The discrepancies can be in the nucleotide sequences, *e.g.* insertions, deletions, or point mutations, or in the amount of a given nucleotide sequence. Methods to determine these discrepancies in sequences are well-known to one of ordinary skill in the art. The “control” nucleic acid target region refers to the sequence or amount of the sequence found in normal cells, *e.g.* cells that are  
10       not diseased as discussed previously.

      The summary of the invention described above is not limiting and other features and advantages of the invention will be apparent from the following detailed description of the invention, and from the claims.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A-1EE shows the nucleotide sequences for human protein kinases oriented in a 5' to 3' direction (SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57).

15

Figures 2A-2L show the amino acid sequences for the human protein kinases encoded by SEQ ID No. 1-57 in the direction of translation (SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114). Some of the sequences encode predicted stop codons within the coding region, indicated by an 'x.'

25

### DETAILED DESCRIPTION OF THE INVENTION

The invention provides, *inter alia*, protein kinase and kinase-like genes, as well as fragments thereof, which have been identified in genomic databases. In part, the invention provides nucleic acid molecules that are capable of encoding polypeptides having a kinase or kinase-like activity. By reference to Tables 1 through 8, below, genes of the invention can be better understood. The invention additionally provides a number of different embodiments, such as those described below.

#### 10 Nucleic Acids

Associations of chromosomal localizations for mapped genes with amplicons implicated in cancer are based on literature searches (PubMed <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>), OMIM searches (Online Mendelian Inheritance in Man, <http://www.ncbi.nlm.nih.gov/Omim/searchomim.html>) and the comprehensive database of cancer amplicons maintained by Knuutila, et al. (Knuutila, et al., DNA copy number amplifications in human neoplasms. Review of comparative genomic hybridization studies. Am J Pathol 152:1107-1123, 1998. <http://www.helsinki.fi/~jgl/www/CMG.html>). For many of the mapped genes, the cytogenetic region from Knuutila is listed followed by the number of cases with documented amplification and the total number of cases studied. Thus for SGK187, the entry “non-small cell lung cancer (12q24.1-24.3; 2/50)” means that the chromosomal position has been associated with non-small cell lung cancer, at position 12q24.1-24.3, which encompasses the SGK087’s position, and the amplification has been noted in 2 of the 50 samples studied.

For single nucleotide polymorphisms, an accession number (for example, ss1581624 for SGK187) is given if the SNP is documented in dbSNP (the database of single nucleotide polymorphisms) maintained at NCBI (<http://www.ncbi.nlm.nih.gov/SNP/index.html>). The accession number for SNP can be used to retrieve the full SNP-containing sequence from this site. Candidate SNPs without a dbSNP accession number were identified by inspection of Blastn

outputs of the patent sequences vs cDNA and genomic databases as indicated, for example, in Tables 9 and 10, provided in Example 1.

Nucleic Acid Probes, Methods, and Kits for Detection of Kinases

5           The invention additionally provides nucleic acid probes and uses therefor. A nucleic acid probe of the present invention may be used to probe an appropriate chromosomal or cDNA library by usual hybridization methods to obtain other nucleic acid molecules of the present invention. A chromosomal DNA or cDNA library may be prepared from appropriate cells according to recognized methods in the art (*cf.* "Molecular Cloning: A Laboratory Manual",  
10       second edition, Cold Spring Harbor Laboratory, Sambrook, Fritsch, & Maniatis, eds., 1989).

          In the alternative, chemical synthesis can be carried out in order to obtain nucleic acid probes having nucleotide sequences which correspond to N-terminal and C-terminal portions of the amino acid sequence of the polypeptide of interest. The synthesized nucleic acid probes may be used as primers in a polymerase chain reaction (PCR) carried out in accordance with  
15       recognized PCR techniques, essentially according to PCR Protocols, "A Guide to Methods and Applications", Academic Press, Michael, *et al.*, eds., 1990, utilizing the appropriate chromosomal or cDNA library to obtain the fragment of the present invention.

          One skilled in the art can readily design such probes based on the sequence disclosed herein using methods of computer alignment and sequence analysis known in the art ("Molecular  
20       Cloning: A Laboratory Manual", 1989, *supra*). The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes may be visualized using known methods.

          The nucleic acid probes of the present invention include RNA, as well as DNA probes,  
25       such probes being generated using techniques known in the art. The nucleic acid probe may be immobilized on a solid support. Examples of such solid supports include, but are not limited to, plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, and acrylic resins, such as polyacrylamide and latex beads. Techniques for coupling nucleic acid probes to such solid supports are well known in the art.

The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid  
5 extracts of cells are well known in the art and can be readily adapted in order to obtain a sample which is compatible with the method utilized.

One method of detecting the presence of nucleic acids of the invention in a sample comprises (a) contacting said sample with the above-described nucleic acid probe under conditions such that hybridization occurs, and (b) detecting the presence of said probe bound to  
10 said nucleic acid molecule. One skilled in the art would select the nucleic acid probe according to techniques known in the art as described above. Samples to be tested include but should not be limited to RNA samples of human tissue.

A kit for detecting the presence of nucleic acids of the invention in a sample comprises at least one container means having disposed therein the above-described nucleic acid probe. The  
15 kit may further comprise other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound nucleic acid probe. Examples of detection reagents include, but are not limited to radiolabelled probes, enzymatic labeled probes (horseradish peroxidase, alkaline phosphatase), and affinity labeled probes (biotin, avidin, or streptavidin). Preferably, the kit further comprises instructions for use.

20 In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion  
25 from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the probe or primers used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, and the like), and containers which contain the reagents used to detect the hybridized probe, bound antibody, amplified product, or the like. One skilled in the art will readily recognize that the nucleic acid



probes described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

## CATEGORIZATION OF THE POLYPEPTIDES ACCORDING TO THE INVENTION

For a number of protein kinases of the invention, there is provided a classification of the protein class and family to which it belongs, a summary of non-catalytic protein motifs, as well as a chromosomal location. This information is useful in determining function, regulation and/or therapeutic utility for each of the proteins. Amplification of chromosomal region can be associated with various cancers. For amplicons discussed in this application, the source of information was Knuutila, et al (Knuutila S, Björkqvist A-M, Autio K, Tarkkanen M, Wolf M, Monni O, Szymanska J, Larramendy ML, Tapper J, Pere H, El-Rifai W, Hemmer S, Wasenius V-M, Vidgren V & Zhu Y: DNA copy number amplifications in human neoplasms. Review of comparative genomic hybridization studies. Am J Pathol 152:1107-1123, 1998. [http://www.helsinki.fi/~lgl\\_www/CMG.html](http://www.helsinki.fi/~lgl_www/CMG.html)).

The kinase classification and protein domains often reflect pathways, cellular roles, or mechanisms of up- or down-stream regulation. Also disease-relevant genes often occur in families of related genes. For example, if one member of a kinase family functions as an oncogene, a tumor suppressor, or has been found to be disrupted in an immune, neurologic, cardiovascular, or metabolic disorder, frequently other family members may play a related role.

The expression analysis organizes kinases into groups that are transcriptionally upregulated in tumors and those that are more restricted to specific tumor types such as melanoma or prostate. This analysis also identifies genes that are regulated in a cell cycle dependent manner, and are therefore likely to be involved in maintaining cell cycle checkpoints, entry, progression, or exit from mitosis, oversee DNA repair, or are involved in cell proliferation and genome stability. Expression data also can identify genes expressed in endothelial sources or other tissues that suggest a role in angiogenesis, thereby implicating them as targets for control of diseases that have an angiogenic component, such as cancer, endometriosis, retinopathy and macular degeneration, and various ischemic or vascular pathologies. A proteins' role in cell survival can also be suggested based on restricted expression in cells subjected to external stress such as oxidative damage, hypoxia, drugs such as cisplatin, or irradiation. Metastases-associated genes can be implicated when expression is restricted to invading regions of a tumor,

or is only seen in local or distant metastases compared to the primary tumor, or when a gene is upregulated during cell culture models of invasion, migration, or motility.

Chromosomal location can identify candidate targets for a tumor amplicon or a tumor-suppressor locus. Summaries of prevalent tumor amplicons are available in the literature, and  
5 can identify tumor types to experimentally be confirmed to contain amplified copies of a kinase gene which localizes to an adjacent region.

As described herein, the polypeptides of the present invention can be classified, for example, among ten different groups. The salient features related to the biological and clinical implications of these different groups are described hereafter in more general terms.

10 A more specific characterization of the polypeptides of the invention, including potential biological and clinical implications, is provided, e.g., in EXAMPLES 2a and 2b.

#### *CLASSIFICATION OF POLYPEPTIDES EXHIBITING KINASE ACTIVITY*

15 The following information also is referenced, for example, at Tables 1 and 2.

##### **AGC Group**

Family members are described that belong to the AGC group of protein kinases. The AGC group of protein kinases includes as its major prototypes protein kinase C (PKC), cAMP-  
20 dependent protein kinases (PKA), the G protein-coupled receptor kinases (ARK and rhodopsin kinase (GRK1)) as well as p70S6K and AKT.

Potential biological and clinical implications of the novel AGC group protein kinases are described in Example 2e. Novel AGC group kinases include: SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, and SEQ ID NO:64.

25

##### **CAMK Group**

Family members are described that belong to the CAMK group of protein kinases. The CAMK group of protein kinases includes as its major prototypes the calmodulin-dependent protein kinases, elongation factor-2 kinases, phosphorylase kinase and the Snf1 and cAMP-  
30 dependent family of protein kinases.

Potential biological and clinical implications of the novel CAMK group of protein kinases are described in Example 2e. Novel CAMK group of protein kinases include: SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, and SEQ ID NO:78.

5

#### **Casein kinase group**

Family members are described that belong to the casein kinase (CKI) group of protein kinases. The Casein kinase (CK) group of protein kinases includes as its major prototypes casein kinaseI (CKI) and casein kinaseII (CKII). Both CKI and CKII are ubiquitous, constitutively-  
10 active, second-messenger-independent kinases, highly conserved enzymes that exist in multiple isoforms. CKI functions in vesicular trafficking, DNA repair, cell cycle progression and cytokinesis (Cell Signal 1998 Nov;10(10):699-711). CK2 functions in cell cycle progression in non-neural cells. CK2 has also been implicated in multiple signaling pathways in normal and disease states of the mammalian nervous systems (Prog Neurobiol 2000 Feb;60(3):211-46 ).

15 Potential biological and clinical implications of the novel casein kinase group of protein kinases are described in Example 2e. Novel casein kinase protein kinases include: SEQ ID NO: 79, and SEQ ID NO:80.

#### **CMGC group**

20 Family members are described that belong to the cyclin-dependent kinase (CDK) group of protein kinases. The CMGC group of protein kinases includes as its major prototypes the cyclin-dependent protein kinases, as well as the MAPK kinases family, the GSK family and the CLK family of kinases.

Potential biological and clinical implications of the novel CMGC group of protein  
25 kinases are described in Example 2e. Novel CMGC protein kinases include: SEQ ID NO:81, SEQ ID NO:82, and SEQ ID NO: 83.

#### **Microbial PK group**

Family members are described that belong to the microbial group of protein kinases. This group is defined, for example, by the protein kinases that include ABC1, RI01, YGR262, all of

which have been initially identified from microbial genome sequencing projects (Proc Natl Acad Sci U S A 1999 Nov 23;96(24):13603-10).

Potential biological and clinical implications of the novel microbial group of protein kinases are described in Example 2e. Novel microbial protein kinases include SEQ ID NO:84,  
5 SEQ ID NO:85, and SEQ ID NO:86.

#### **“Other” group**

Family members are described that belong to the “Other” group of protein kinases. Within this group of protein kinases are members that have recognizable catalytic motifs that are  
10 identifiable by a hidden Markov model analysis, but fail to cluster with other protein kinases on the basis of their amino acid sequence homology over the catalytic region.

Potential biological and clinical implications of the novel protein kinases belonging to the Other group are described in Example 2e. Novel “Other” protein kinases include: SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID  
15 NO:93 , SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, and SEQ ID NO:101.

#### **The STE Group**

Family members are described that belong to the STE group of protein kinases. The STE  
20 group of protein kinases includes, as its major prototypes, the NEK kinases, as well as the STE11 and STE20 family of sterile protein kinases.

Potential biological and clinical implications of the novel protein kinases belonging to the STE group are described in Example 2e. Novel STE protein kinases include: SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ  
25 ID NO:108, and SEQ ID NO:109.

#### **The TK group**

Family members are described that belong to the tyrosine kinase (TK) group of protein kinases. The TK group of protein kinases includes as its major prototypes the cytoplasmic and  
30 receptor families of protein kinases.

One family within this group of kinases is the “Eph” family. The Eph family, which is the largest sub-family of receptor tyrosine kinases in the human genome, has a stereotyped structure consisting of an N-terminal globular domain involved in ligand binding, two Type III fibronectin-like domains which contribute to receptor dimerization, a transmembrane domain, and an intracellular tyrosine kinase domain. The Eph family is composed of two subfamilies: the EphA receptors which generally bind to members of the GPI-linked Ephrin A family of ligands, and the EphB receptors which generally bind to the transmembrane Ephrin B family of ligands. Based on sequence similarity comparisons, EphA9 – to which the polypeptide represented by SEQ ID NO:110 belongs, is a member of the EphA subfamily of receptors.

Investigation of the Eph family of receptors indicate their involvement in a wide variety of cellular processes. Activation of Eph receptors can lead to changes in intracellular signaling, cell adhesion, cytoskeleton effects, and synaptic remodeling. These Eph-dependent cellular effects in turn contribute to changes in tissue functions such as border formation, pattern formation, cell migration, neurogenesis, angiogenesis, and long term potentiation, among others. As a member of the Eph family, we expect that EphA9 will be involved in many of these functions as well.

Expression data for EphA9 indicate that it is expressed most prominently in the human central nervous system, the digestive system (especially in the colon and rectum) and the testes. Therefore, EphA9 may be involved in organization and function of the digestive tract, including the colon, and could contribute colorectal tumorigenesis and other disorders of the digestive tract.

EphA9 found in the nervous system could be involved in synaptogenesis, neuronal development and regeneration, axon outgrowth, and synaptic transmission. Therefore EphA9 might be important in neuronal survival and regeneration after injury, in long-term potentiation and memory formation, and in disorders of synaptic transmission such as epilepsy, depression, Parkinson’s disease, and Alzheimer’s disease.

Several Eph family receptors previously have been shown to be critical to several aspects of angiogenesis, such as remodeling, branching, sprouting and pruning of new blood vessels. EphA9, as an additional member of this family, also may be critical for aspects of angiogenesis.

Thus, EphA9 may be relevant for a number diseases, including solid tumors, rheumatoid arthritis, and cardiovascular diseases.

Potential biological and clinical implications of the novel protein kinases belonging to the TK group are described in Example 2e. Novel TK protein kinases include: SEQ ID NO:100, and  
5 SEQ ID NO:111.

#### *CLASSIFICATION OF POLYPEPTIDES EXHIBITING KINASE-LIKE ACTIVITY*

Two new family members are described that belong to the protein kinase (PK)-like insert  
10 “super family” of protein kinases. The PK-like superfamily of protein kinases includes the diacyl glycerol kinases (DGK) and the guanylate cyclases (GCyc), as described in the EXAMPLES.

##### **Guanylate Cyclases (GCyc) Group**

Guanylate cyclases are ubiquitous enzymes that convert GTP to cGMP and exist as  
15 membrane-bound and soluble isoforms. A diverse range of agonist that include peptide hormones, bacterial toxins as well as intracellular molecules such as calcium and cAMP regulate the enzymatic activity of guanylate cyclases. Stimulation of guanylate kinases modulates multiple downstream enzymes including cGMP-dependent protein kinases, cGMP-regulated phosphodiesterases, and cyclic nucleotide-gated ion channels. The modulation of cGMP levels  
20 by guanylate cyclases contributes to the regulation of vascular smooth muscle motility, intestinal fluid and electrolyte homeostasis, and retinal phototransduction (Pharmacol Rev 2000 Sep;52(3):375-414). As potential novel members of the guanylate cyclase family, disruptions in the signaling pathways in which SGK007 and SGK050 participate may alter cGMP homeostasis with pathophysiological implications.

25

##### **Diacyl Glycerol kinase (DGK) Group**

A diacyl glycerol kinase phosphorylates the second messenger molecule diacyl glycerol leading to the formation of phosphatidic acid. Nine mammalian DGK isozymes have been  
30 described. The catalytic domain of a DGK usually is flanked by protein-protein interaction

domains such as zinc fingers, pleckstrin homology domains and ankyrin repeats, as well as calcium-binding EF-hand structures. DGK's can be associated with the plasma membrane, nucleus and cytoskeleton. Experimental evidence supports the proposition that DGK's are translocated to and from these cellular compartments in response to agonists. At these  
5 intracellular locations, DGK's are able to modulate lipid metabolism and PKC activation, thereby triggering effector functions related to cell cycle progression and differentiation (*Int. J. Biochem. Cell Biol.* 1997, (10):1139-43, *J. Biol. Chem.* 1999, 274(17):11447-50.)

#### 10 **SGK093 - The Wnk family of serine/threonine kinases**

Wnk3 is a member of a subfamily of serine/threonine kinases which includes a described prototype, Wnk1, isolated from rat. This family is characterized by an N-terminal catalytic domain with several unique sequence features, most notably a change of the invariant lysine in kinase subdomain II to a cysteine, coupled with a change of the third conserved glycine residue  
15 in subdomain I into a lysine. The resulting enzyme appears to maintain catalytic activity due to this concomitant switch. Wnk3 conserves both of these catalytic changes and therefore is predicted to maintain catalytic activity. The long C-terminal portion of the wnk3 includes many protein interaction domains such as SH3 binding sites and coiled coil regions.

The wnk family catalytic domain shows the highest similarity to two families of  
20 serine/threonine kinases: The MEKK-like kinases and the Ste20-like kinases. Both of these families can regulate enzymes in various MAPK signaling cascades, which are critical for many cellular processes such as mitogenesis, differentiation, cell survival, and stress response. The Ste20 kinases are also involved in regulation of the ras/rac/rho/cdc42 pathways and subsequent downstream effects on cytoskeleton.

25 Wnk3 shows high expression in human kidney, in kidney carcinoma cell lines, in prostate, prostate cell lines, and prostate tumor bone metastases, in colorectal tissue and tumor cell lines, and in human leukemia cells. Therefore wnk3 may be involved in the normal homeostasis and functioning of the human kidney, prostate, and digestive system, and may be involved in tumorigenesis which arises from these three tissues. High expression in human  
30 leukemia cell lines indicates a possible role in the development of that disease as well.



*THERAPEUTIC METHODS ACCORDING TO THE INVENTION:***Diagnostics:**

5

The invention provides methods for detecting a polypeptide in a sample as a diagnostic tool for diseases or disorders, wherein the method comprises the steps of: (a) contacting the sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid target region of a polypeptide selected from the group consisting of SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, said probe comprising the nucleic acid sequence encoding the polypeptide, fragments thereof, and the complements of the sequences and fragments; and (b) detecting the presence or amount of the probe:target region hybrid as an indication of the disease.

In preferred embodiments of the invention, the disease or disorder is selected from the group consisting of rheumatoid arthritis, atherosclerosis, autoimmune disorders, organ transplantation, myocardial infarction, cardiomyopathies, stroke, renal failure, oxidative stress-related neurodegenerative disorders, metabolic disorder including diabetes, reproductive disorders including infertility, and cancer.

Hybridization conditions should be such that hybridization occurs only with the genes in the presence of other nucleic acid molecules. Under stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having 1 or 2 mismatches out of 20 contiguous nucleotides. Such conditions are defined supra.

The diseases for which detection of genes in a sample could be diagnostic include diseases in which nucleic acid (DNA and/or RNA) is amplified in comparison to normal cells. By "amplification" is meant increased numbers of DNA or RNA in a cell compared with normal cells.

5 "Amplification" as it refers to RNA can be the detectable presence of RNA in cells, since in some normal cells there is no basal expression of RNA. In other normal cells, a basal level of expression exists, therefore in these cases amplification is the detection of at least 1-2-fold, and preferably more, compared to the basal level.

The diseases that could be diagnosed by detection of nucleic acid in a sample preferably  
10 include cancers. The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to  
15 obtain a sample that is compatible with the method utilized.

#### **Antibodies, Hybridomas, Methods of Use and Kits for Detection of Kinases**

The present invention relates to an antibody having binding affinity to a kinase of the invention.  
20 The polypeptide may have the amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79,  
25 SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID  
30 NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, or a functional derivative thereof, or at least 9.

contiguous amino acids thereof (preferably, at least 20, 30, 35, or 40 contiguous amino acids thereof).

The present invention also relates to an antibody having specific binding affinity to a kinase of the invention. Such an antibody may be isolated by comparing its binding affinity to a kinase of the invention with its binding affinity to other polypeptides. Those which bind selectively to a kinase of the invention would be chosen for use in methods requiring a distinction between a kinase of the invention and other polypeptides. Such methods could include, but should not be limited to, the analysis of altered kinase expression in tissue containing other polypeptides.

The kinases of the present invention can be used in a variety of procedures and methods, such as for the generation of antibodies, for use in identifying pharmaceutical compositions, and for studying DNA/protein interaction.

The kinases of the present invention can be used to produce antibodies or hybridomas. One skilled in the art will recognize that if an antibody is desired, such a peptide could be generated as described herein and used as an immunogen. The antibodies of the present invention include monoclonal and polyclonal antibodies, as well fragments of these antibodies, and humanized forms. Humanized forms of the antibodies of the present invention may be generated using one of the procedures known in the art such as chimerization or CDR grafting.

The present invention also relates to a hybridoma which produces the above-described monoclonal antibody, or binding fragment thereof. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing monoclonal antibodies and hybridomas are well known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1984; St. Groth *et al.*, *J. Immunol. Methods* 35:1-21, 1980). Any animal (mouse, rabbit, and the like) which is known to produce antibodies can be immunized with the selected polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of polypeptide used for immunization will vary based on the animal which is immunized, the antigenicity of the polypeptide and the site of injection.

The polypeptide may be modified or administered in an adjuvant in order to increase the peptide antigenicity. Methods of increasing the antigenicity of a polypeptide are well known in the art. Such procedures include coupling the antigen with a heterologous protein (such as globulin or  $\beta$ -galactosidase) or through the inclusion of an adjuvant during immunization.

5 For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot  
10 analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175:109-124, 1988). Hybridomas secreting the desired antibodies are cloned and the class and subclass are determined using procedures known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", *supra*, 1984).

For polyclonal antibodies, antibody-containing antisera is isolated from the immunized  
15 animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The above-described antibodies may be detectably labeled. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, and the like), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, and the like) fluorescent labels (such as FITC or rhodamine, and the like),  
20 paramagnetic atoms, and the like. Procedures for accomplishing such labeling are well-known in the art, for example, *see* Stemberger *et al.*, *J. Histochem. Cytochem.* 18:315, 1970; Bayer *et al.*, *Meth. Enzym.* 62:308, 1979; Engval *et al.*, *Immunol.* 109:129, 1972; Goding, *J. Immunol. Meth.* 13:215, 1976. The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues which express a specific peptide.

25 The above-described antibodies may also be immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England,  
30 Chapter 10, 1986; Jacoby *et al.*, *Meth. Enzym.* 34, Academic Press, N.Y., 1974). The

immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as in immunochromatography.

Furthermore, one skilled in the art can readily adapt currently available procedures, as well as the techniques, methods and kits disclosed herein with regard to antibodies, to generate peptides capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides (Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides", In Synthetic Peptides, A User's Guide, W.H. Freeman, NY, pp. 289-307, 1992; Kaspczak *et al.*, *Biochemistry* 28:9230-9238, 1989).

Anti-peptide peptides can be generated by replacing the basic amino acid residues found in the peptide sequences of the kinases of the invention with acidic residues, while maintaining hydrophobic and uncharged polar groups. For example, lysine, arginine, and/or histidine residues are replaced with aspartic acid or glutamic acid and glutamic acid residues are replaced by lysine, arginine or histidine.

The present invention also encompasses a method of detecting a kinase polypeptide in a sample, comprising: (a) contacting the sample with an above-described antibody, under conditions such that immunocomplexes form, and (b) detecting the presence of said antibody bound to the polypeptide. In detail, the methods comprise incubating a test sample with one or more of the antibodies of the present invention and assaying whether the antibody binds to the test sample. Altered levels of a kinase of the invention in a sample as compared to normal levels may indicate disease.

Conditions for incubating an antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the antibody used in the assay. One skilled in the art will recognize that any one of the commonly available immunological assay formats (such as radioimmunoassays, enzyme-linked immunosorbent assays, diffusion-based Ouchterlony, or rocket immunofluorescent assays) can readily be adapted to employ the antibodies of the present invention. Examples of such assays can be found in Chard ("An Introduction to Radioimmunoassay and Related Techniques" Elsevier Science Publishers, Amsterdam, The Netherlands, 1986), Bullock *et al.* ("Techniques in Immunocytochemistry," Academic Press, Orlando, FL Vol. 1, 1982; Vol. 2, 1983; Vol. 3, 1985), Tijssen ("Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in

Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1985).

The immunological assay test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as blood, serum, plasma, or urine. The test samples used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can readily be adapted in order to obtain a sample which is testable with the system utilized.

A kit contains all the necessary reagents to carry out the previously described methods of detection. The kit may comprise: (i) a first container means containing an above-described antibody, and (ii) second container means containing a conjugate comprising a binding partner of the antibody and a label. In another preferred embodiment, the kit further comprises one or more other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound antibodies.

Examples of detection reagents include, but are not limited to, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the chromophoric, enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. The compartmentalized kit may be as described above for nucleic acid probe kits. One skilled in the art will readily recognize that the antibodies described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

#### **Isolation of Compounds Capable of Interacting with Kinases**

The present invention also relates to a method of detecting a compound capable of binding to a kinase of the invention comprising incubating the compound with a kinase of the invention and detecting the presence of the compound bound to the kinase. The compound may be present within a complex mixture, for example, serum, body fluid, or cell extracts.

The present invention also relates to a method of detecting an agonist or antagonist of kinase activity or kinase binding partner activity comprising incubating cells that produce a kinase of the invention in the presence of a compound and detecting changes in the level of kinase activity or kinase binding partner activity. The compounds thus identified would produce

a change in activity indicative of the presence of the compound. The compound may be present within a complex mixture, for example, serum, body fluid, or cell extracts. Once the compound is identified it can be isolated using techniques well known in the art.

**5 Modulating polypeptide activity:**

The invention additionally provides methods for treating a disease or abnormal condition by administering to a patient in need of such treatment a substance that modulates the activity of a polypeptide selected from the group consisting of SEQ ID NO:58, SEQ ID NO:59, SEQ ID  
10 NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88,  
15 SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114. Preferably,  
20 the disease is selected from the group consisting of rheumatoid arthritis, atherosclerosis, autoimmune disorders, organ transplantation, myocardial infarction, cardiomyopathies, stroke, renal failure, oxidative stress-related neurodegenerative disorders, metabolic and reproductive disorders, and cancer.

Substances useful for treatment of disorders or diseases preferably show positive results  
25 in one or more assays for an activity corresponding to treatment of the disease or disorder in question. Substances that modulate the activity of the polypeptides preferably include, but are not limited to, antisense oligonucleotides and inhibitors of protein kinases.

The term "preventing" refers to decreasing the probability that an organism contracts or develops an abnormal condition.

The term "treating" refers to having a therapeutic effect and at least partially alleviating or abrogating an abnormal condition in the organism.

The term "therapeutic effect" refers to the inhibition or activation factors causing or contributing to the abnormal condition. A therapeutic effect relieves to some extent one or more of the symptoms of the abnormal condition. In reference to the treatment of abnormal conditions, a therapeutic effect can refer to one or more of the following: (a) an increase in the proliferation, growth, and/or differentiation of cells; (b) inhibition (, slowing or stopping) of cell death; (c) inhibition of degeneration; (d) relieving to some extent one or more of the symptoms associated with the abnormal condition; and (e) enhancing the function of the affected population of cells. Compounds demonstrating efficacy against abnormal conditions can be identified as described herein.

The term "abnormal condition" refers to a function in the cells or tissues of an organism that deviates from their normal functions in that organism. An abnormal condition can relate to cell proliferation, cell differentiation or cell survival. An abnormal condition may also include irregularities in cell cycle progression, i.e., irregularities in normal cell cycle progression through mitosis and meiosis.

Abnormal cell proliferative conditions include cancers such as fibrotic and mesangial disorders, abnormal angiogenesis and vasculogenesis, wound healing, psoriasis, diabetes mellitus, and inflammation.

Abnormal differentiation conditions include, but are not limited to, neurodegenerative disorders, slow wound healing rates, and slow tissue grafting healing rates.

Abnormal cell survival conditions may also relate to conditions in which programmed cell death (apoptosis) pathways are activated or abrogated. A number of protein kinases are associated with the apoptosis pathways. Aberrations in the function of any one of the protein kinases could lead to cell immortality or premature cell death.

The term "aberration", in conjunction with the function of a kinase in a signal transduction process, refers to a kinase that is over- or under-expressed in an organism, mutated such that its catalytic activity is lower or higher than wild-type protein kinase activity, mutated such that it can no longer interact with a natural binding partner, is no longer modified by



another protein kinase or protein phosphatase, or no longer interacts with a natural binding partner.

The term "administering" relates to a method of incorporating a compound into cells or tissues of an organism. The abnormal condition can be prevented or treated when the cells or  
5 tissues of the organism exist within the organism or outside of the organism. Cells existing outside the organism can be maintained or grown in cell culture dishes. For cells harbored within the organism, many techniques exist in the art to administer compounds, including (but not limited to) oral, parenteral, dermal, injection, and aerosol applications. For cells outside of  
10 the organism, multiple techniques exist in the art to administer the compounds, including (but not limited to) cell microinjection techniques, transformation techniques and carrier techniques.

The abnormal condition can also be prevented or treated by administering a compound to a group of cells having an aberration in a signal transduction pathway to an organism. The effect of administering a compound on organism function can then be monitored. The organism is preferably a mouse, rat, rabbit, guinea pig or goat, more preferably a monkey or ape, and most  
15 preferably a human.

The present invention also encompasses a method of agonizing (stimulating) or antagonizing kinase associated activity in a mammal comprising administering to said mammal an agonist or antagonist to a kinase of the invention in an amount sufficient to effect said  
20 agonism or antagonism. A method of treating diseases in a mammal with an agonist or antagonist of the activity of one of the kinases of the invention comprising administering the agonist or antagonist to a mammal in an amount sufficient to agonize or antagonize kinase-associated functions is also encompassed in the present application.

In an effort to discover novel treatments for diseases, biomedical researchers and chemists have designed, synthesized, and tested molecules that inhibit the function of protein  
25 kinases. Some small organic molecules form a class of compounds that modulate the function of protein kinases. Examples of molecules that have been reported to inhibit the function of protein kinases include, but are not limited to, bis monocyclic, bicyclic or heterocyclic aryl compounds (PCT WO 92/20642, published November 26, 1992 by Maguire *et al.*), vinylene-azaindole derivatives (PCT WO 94/14808, published July 7, 1994 by Ballinari *et al.*), 1-cyclopropyl-4-  
30 pyridyl-quinolones (U.S. Patent No. 5,330,992), styryl compounds (U.S. Patent No. 5,217,999),

styryl-substituted pyridyl compounds (U.S. Patent No. 5,302,606), certain quinazoline derivatives (EP Application No. 0 566 266 A1), selcoindoles and selenides (PCT WO 94/03427, published February 17, 1994 by Denny *et al.*), tricyclic polyhydroxylic compounds (PCT WO 92/21660, published December 10, 1992 by Dow), and benzylphosphonic acid compounds (PCT  
5 WO 91/15495, published October 17, 1991 by Dow *et al.*).

Compounds that can traverse cell membranes and are resistant to acid hydrolysis are potentially advantageous as therapeutics as they can become highly bioavailable after being administered orally to patients. However, many of these protein kinase inhibitors only weakly inhibit the function of protein kinases. In addition, many inhibit a variety of protein kinases and  
10 will therefore cause multiple side-effects as therapeutics for diseases.

Some indolinone compounds, however, form classes of acid resistant and membrane permeable organic molecules. WO 96/22976 (published August 1, 1996 by Ballinari *et al.*) describes hydrosoluble indolinone compounds that harbor tetralin, naphthalene, quinoline, and indole substituents fused to the oxindole ring. These bicyclic substituents are in turn substituted  
15 with polar moieties including hydroxylated alkyl, phosphate, and ether moieties. U.S. Patent Application Serial Nos. 08/702,232, filed August 23, 1996, entitled "Indolinone Combinatorial Libraries and Related Products and Methods for the Treatment of Disease" by Tang *et al.* (Lyon & Lyon Docket No. 221/187) and 08/485,323, filed June 7, 1995, entitled "Benzyldiene-Z-Indoline Compounds for the Treatment of Disease" by Tang *et al.* (Lyon & Lyon Docket No.  
20 223/298) and International Patent Publications WO 96/40116, published December 19, 1996 by Tang, *et al.*, and WO 96/22976, published August 1, 1996 by Ballinari *et al.*, all of which are incorporated herein by reference in their entirety, including any drawings, figures, or tables, describe indolinone chemical libraries of indolinone compounds harboring other bicyclic moieties as well as monocyclic moieties fused to the oxindole ring. Applications 08/702,232,  
25 filed August 23, 1996, entitled "Indolinone Combinatorial Libraries and Related Products and Methods for the Treatment of Disease" by Tang *et al.* (Lyon & Lyon Docket No. 221/187), 08/485,323, filed June 7, 1995, entitled "Benzyldiene-Z-Indoline Compounds for the Treatment of Disease" by Tang *et al.* (Lyon & Lyon Docket No. 223/298), and WO 96/22976, published August 1, 1996 by Ballinari *et al.* teach methods of indolinone synthesis, methods of testing the

biological activity of indolinone compounds in cells, and inhibition patterns of indolinone derivatives.

Other examples of substances capable of modulating kinase activity include, but are not limited to, tyrphostins, quinazolines, quinoxolines, and quinolines. The quinazolines, tyrphostins, quinolines, and quinoxolines referred to above include well known compounds such as those described in the literature. For example, representative publications describing quinazolines include Barker *et al.*, EPO Publication No. 0 520 722 A1; Jones *et al.*, U.S. Patent No. 4,447,608; Kabbe *et al.*, U.S. Patent No. 4,757,072; Kaul and Vougioukas, U.S. Patent No. 5,316,553; Kreighbaum and Comer, U.S. Patent No. 4,343,940; Pegg and Wardleworth, EPO Publication No. 0 562 734 A1; Barker *et al.*, (1991) *Proc. of Am. Assoc. for Cancer Research* 32:327; Bertino, J.R., (1979) *Cancer Research* 3:293-304; Bertino, J.R., (1979) *Cancer Research* 9(2 part 1):293-304; Curtin *et al.*, (1986) *Br. J. Cancer* 53:361-368; Fernandes *et al.*, (1983) *Cancer Research* 43:1117-1123 ; Ferris *et al. J. Org. Chem.* 44(2):173-178; Fry *et al.*, (1994) *Science* 265:1093-1095; Jackman *et al.*, (1981) *Cancer Research* 51:5579-5586; Jones *et al. J. Med. Chem.* 29(6):1114-1118; Lee and Skibo, (1987) *Biochemistry* 26(23):7355-7362; Lemus *et al.*, (1989) *J. Org. Chem.* 54:3511-3518; Ley and Seng, (1975) *Synthesis* 1975:415-522; Maxwell *et al.*, (1991) *Magnetic Resonance in Medicine* 17:189-196 ; Mini *et al.*, (1985) *Cancer Research* 45:325-330; Phillips and Castle, J. (1980) *Heterocyclic Chem.* 17(19):1489-1596; Reece *et al.*, (1977) *Cancer Research* 47(11):2996-2999; Sculier *et al.*, (1986) *Cancer Immunol. and Immunother.* 23, A65; Sikora *et al.*, (1984) *Cancer Letters* 23:289-295; Sikora *et al.*, (1988) *Analytical Biochem.* 172:344-355; all of which are incorporated herein by reference in their entirety, including any drawings.

Quinoxaline is described in Kaul and Vougioukas, U.S. Patent No. 5,316,553, incorporated herein by reference in its entirety, including any drawings.

Quinolines are described in Dolle *et al.*, (1994) *J. Med. Chem.* 37:2627-2629; MaGuire, J. (1994) *Med. Chem.* 37:2129-2131; Burke *et al.*, (1993) *J. Med. Chem.* 36:425-432 ; and Burke *et al.* (1992) *BioOrganic Med. Chem. Letters* 2:1771-1774, all of which are incorporated by reference in their entirety, including any drawings.

Tyrphostins are described in Allen *et al.*, (1993) *Clin. Exp. Immunol.* 91:141-156; Anafi *et al.*, (1993) *Blood* 82:12, 3524-3529; Baker *et al.*, (1992) *J. Cell Sci.* 102:543-555; Bilder *et al.*,

(1991) *Amer. Physiol. Soc.* pp. 6363-6143:C721-C730; Brunton *et al.*, (1992) *Proceedings of Amer. Assoc. Cancer Rsch.* 33:558; Bryckaert *et al.*, (1992) *Exp. Cell Research* 199:255-261; Dong *et al.*, (1993) *J. Leukocyte Biology* 53:53-60; Dong *et al.*, (1993) *J. Immunol.* 151(5):2717-2724; Gazit *et al.*, (1989) *J. Med. Chem.* 32, 2344-2352; Gazit *et al.*, (1993) *J. Med. Chem.* 36:3556-3564; Kaur *et al.*, (1994) *Anti-Cancer Drugs* 5:213-222; King *et al.*, (1991) *Biochem. J.* 275:413-418; Kuo *et al.*, (1993) *Cancer Letters* 74:197-202; Levitzki, A., (1992) *The FASEB J.* 6:3275-3282; Lyall *et al.*, (1989) *J. Biol. Chem.* 264:14503-14509; Peterson *et al.*, (1993) *The Prostate* 22:335-345; Pillemer *et al.*, (1992) *Int. J. Cancer* 50:80-85; Posner *et al.*, (1993) *Molecular Pharmacology* 45:673-683; Rendu *et al.*, (1992) *Biol. Pharmacology* 44(5):881-888; Sauro and Thomas, (1993) *Life Sciences* 53:371-376; Sauro and Thomas, (1993) *J. Pharm. and Experimental Therapeutics* 267(3):119-1125; Wolbring *et al.*, (1994) *J. Biol. Chem.* 269(36):22470-22472; and Yoneda *et al.*, (1991) *Cancer Research* 51:4430-4435; all of which are incorporated herein by reference in their entirety, including any drawings.

Other compounds that could be used as modulators include oxindolinones such as those described in U.S. patent application Serial No. 08/702,232 filed August 23, 1996, incorporated herein by reference in its entirety, including any drawings.

#### RECOMBINANT DNA TECHNOLOGY:

##### **DNA Constructs Comprising a Kinase Nucleic Acid Molecule and Cells Containing These Constructs:**

The present invention also relates to a recombinant DNA molecule comprising, 5' to 3', a promoter effective to initiate transcription in a host cell and the above-described nucleic acid molecules. In addition, the present invention relates to a recombinant DNA molecule comprising a vector and an above-described nucleic acid molecule. The present invention also relates to a nucleic acid molecule comprising a transcriptional region functional in a cell, a sequence complementary to an RNA sequence encoding an amino acid sequence corresponding to the above-described polypeptide, and a transcriptional termination region functional in said cell. The above-described molecules may be isolated and/or purified DNA molecules.

The present invention also relates to a cell or organism that contains an above-described nucleic acid molecule and thereby is capable of expressing a polypeptide. The polypeptide may be purified from cells which have been altered to express the polypeptide. A cell is said to be "altered to express a desired polypeptide" when the cell, through genetic manipulation, is made to produce a protein which it normally does not produce or which the cell normally produces at lower levels. One skilled in the art can readily adapt procedures for introducing and expressing either genomic, cDNA, or synthetic sequences into either eukaryotic or prokaryotic cells.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains nucleotide sequences which contain transcriptional and translational regulatory information and such sequences are "operably linked" to nucleotide sequences which encode the polypeptide. An operable linkage is a linkage in which the regulatory DNA sequences and the DNA sequence sought to be expressed are connected in such a way as to permit gene sequence expression. The precise nature of the regulatory regions needed for gene sequence expression may vary from organism to organism, but shall in general include a promoter region which, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

If desired, the non-coding region 3' to the sequence encoding a kinase of the invention may be obtained by the above-described methods. This region may be retained for its transcriptional termination regulatory sequences, such as termination and polyadenylation. Thus, by retaining the 3'-region naturally contiguous to the DNA sequence encoding a kinase of the invention, the transcriptional termination signals may be provided. Where the transcriptional termination signals are not satisfactorily functional in the expression host cell, then a 3' region functional in the host cell may be substituted.

Two DNA sequences (such as a promoter region sequence and a sequence encoding a kinase of the invention) are said to be operably linked if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region sequence to direct the transcription of a gene sequence

encoding a kinase of the invention, or (3) interfere with the ability of the gene sequence of a kinase of the invention to be transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a DNA sequence if the promoter were capable of effecting transcription of that DNA sequence. Thus, to express a gene encoding a kinase of the invention,  
5 transcriptional and translational signals recognized by an appropriate host are necessary.

The present invention encompasses the expression of a gene encoding a kinase of the invention (or a functional derivative thereof) in either prokaryotic or eukaryotic cells. Prokaryotic hosts are, generally, very efficient and convenient for the production of recombinant proteins and are, therefore, one type of preferred expression system for kinases of the invention.  
10 Prokaryotes most frequently are represented by various strains of *E. coli*. However, other microbial strains may also be used, including other bacterial strains.

In prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host may be used. Examples of suitable plasmid vectors may include pBR322, pUC118, pUC119 and the like; suitable phage or  
15 bacteriophage vectors may include  $\lambda$ gt10,  $\lambda$ gt11 and the like; and suitable virus vectors may include pMAM-neo, pKRC and the like. Preferably, the selected vector of the present invention has the capacity to replicate in the selected host cell.

Recognized prokaryotic hosts include bacteria such as *E. coli*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia*, and the like. However, under such conditions, the  
20 polypeptide will not be glycosylated. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

To express a kinase of the invention (or a functional derivative thereof) in a prokaryotic cell, it is necessary to operably link the sequence encoding the kinase of the invention to a functional prokaryotic promoter. Such promoters may be either constitutive or, more preferably, regulatable (*i.e.*, inducible or derepressible). Examples of constitutive promoters include the *int*  
25 promoter of bacteriophage  $\lambda$ , the *bla* promoter of the  $\beta$ -lactamase gene sequence of pBR322, and the *cat* promoter of the chloramphenicol acetyl transferase gene sequence of pPR325, and the like. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage  $\lambda$  ( $P_L$  and  $P_R$ ), the *trp*,  *$\lambda$ recA*, *acZ*,  *$\lambda$ acI*, and *gal* promoters of *E. coli*, the  $\alpha$ -  
30 amylase (Ulmanen *et al.*, *J. Bacteriol.* 162:176-182, 1985) and the  $\zeta$ -28-specific promoters of *B.*

*subtilis* (Gilman *et al.*, *Gene Sequence* 32:11-20, 1984), the promoters of the bacteriophages of *Bacillus* (Gryczan, in: *The Molecular Biology of the Bacilli*, Academic Press, Inc., NY, 1982), and *Streptomyces* promoters (Ward *et al.*, *Mol. Gen. Genet.* 203:468-478, 1986). Prokaryotic promoters are reviewed by Glick (*Ind. Microbiot.* 1:277-282, 1987), Cenatiempo (*Biochimie* 5 68:505-516, 1986), and Gottesman (*Ann. Rev. Genet.* 18:415-442, 1984).

Proper expression in a prokaryotic cell also requires the presence of a ribosome-binding site upstream of the gene sequence-encoding sequence. Such ribosome-binding sites are disclosed, for example, by Gold *et al.* (*Ann. Rev. Microbiol.* 35:365-404, 1981). The selection of control sequences, expression vectors, transformation methods, and the like, are dependent on 10 the type of host cell used to express the gene. As used herein, "cell", "cell line", and "cell culture" may be used interchangeably and all such designations include progeny. Thus, the words "transformants" or "transformed cells" include the primary subject cell and cultures derived therefrom, without regard to the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent 15 mutations. However, as defined, mutant progeny have the same functionality as that of the originally transformed cell.

Host cells which may be used in the expression systems of the present invention are not strictly limited, provided that they are suitable for use in the expression of the kinase polypeptide of interest. Suitable hosts may often include eukaryotic cells. Preferred eukaryotic hosts 20 include, for example, yeast, fungi, insect cells, mammalian cells either *in vivo*, or in tissue culture. Mammalian cells which may be useful as hosts include HeLa cells, cells of fibroblast origin such as VERO or CHO-K1, or cells of lymphoid origin and their derivatives. Preferred mammalian host cells include SP2/0 and J558L, as well as neuroblastoma cell lines such as IMR 332, which may provide better capacities for correct post-translational processing.

25 In addition, plant cells are also available as hosts, and control sequences compatible with plant cells are available, such as the cauliflower mosaic virus 35S and 19S, and nopaline synthase promoter and polyadenylation signal sequences. Another preferred host is an insect cell, for example the *Drosophila* larvae. Using insect cells as hosts, the *Drosophila* alcohol dehydrogenase promoter can be used (Rubin, *Science* 240:1453-1459, 1988). Alternatively, 30 baculovirus vectors can be engineered to express large amounts of kinases of the invention in

insect cells (Jasny, *Science* 238:1653, 1987; Miller *et al.*, in: *Genetic Engineering*, Vol. 8, Plenum, Setlow *et al.*, eds., pp. 277-297, 1986).

Any of a series of yeast expression systems can be utilized which incorporate promoter and termination elements from the actively expressed sequences coding for glycolytic enzymes that are produced in large quantities when yeast are grown in mediums rich in glucose. Known glycolytic gene sequences can also provide very efficient transcriptional control signals. Yeast provides substantial advantages in that it can also carry out post-translational modifications. A number of recombinant DNA strategies exist utilizing strong promoter sequences and high copy number plasmids which can be utilized for production of the desired proteins in yeast. Yeast recognizes leader sequences on cloned mammalian genes and secretes peptides bearing leader sequences (*i.e.*, pre-peptides). Several possible vector systems are available for the expression of kinases of the invention in a mammalian host.

A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals may be derived from viral sources, such as adenovirus, bovine papilloma virus, cytomegalovirus, simian virus, or the like, where the regulatory signals are associated with a particular gene sequence which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, and the like, may be employed. Transcriptional initiation regulatory signals may be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical (such as metabolite) regulation.

Expression of kinases of the invention in eukaryotic hosts requires the use of eukaryotic regulatory regions. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis. Preferred eukaryotic promoters include, for example, the promoter of the mouse metallothionein I gene sequence (Hamer *et al.*, *J. Mol. Appl. Gen.* 1:273-288, 1982); the TK promoter of Herpes virus (McKnight, *Cell* 31:355-365, 1982); the SV40 early promoter (Benoist *et al.*, *Nature* (London) 290:304-31, 1981); and the yeast gal4 gene sequence promoter (Johnston *et al.*, *Proc. Natl. Acad. Sci. (USA)* 79:6971-6975, 1982; Silver *et al.*, *Proc. Natl. Acad. Sci. (USA)* 81:5951-5955, 1984).



Translation of eukaryotic mRNA is initiated at the codon which encodes the first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a DNA sequence which encodes a kinase of the invention (or a functional derivative thereof) does not contain any intervening codons which are capable of encoding a methionine (*i.e.*, AUG). The presence of such codons results either in the formation of a fusion protein (if the AUG codon is in the same reading frame as the kinase of the invention coding sequence) or a frame-shift mutation (if the AUG codon is not in the same reading frame as the kinase of the invention coding sequence).

A nucleic acid molecule encoding a kinase of the invention and an operably linked promoter may be introduced into a recipient prokaryotic or eukaryotic cell either as a nonreplicating DNA or RNA molecule, which may either be a linear molecule or, more preferably, a closed covalent circular molecule. Since such molecules are incapable of autonomous replication, the expression of the gene may occur through the transient expression of the introduced sequence. Alternatively, permanent expression may occur through the integration of the introduced DNA sequence into the host chromosome.

A vector may be employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, *e.g.*, antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection. Additional elements may also be needed for optimal synthesis of mRNA. These elements may include splice signals, as well as transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama (*Mol. Cell. Biol.* 3:280-289, 1983).

The introduced nucleic acid molecule can be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and

selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Preferred prokaryotic vectors include plasmids such as those capable of replication in *E. coli* (such as, for example, pBR322, ColEI, pSC101, pACYC 184,  $\pi$ VX; "Molecular Cloning: A Laboratory Manual", 1989, *supra*). *Bacillus* plasmids include pC194, pC221, pT127, and the like (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, NY, pp. 307-329, 1982). Suitable *Streptomyces* plasmids include p1J101 (Kendall *et al.*, *J. Bacteriol.* 169:4177-4183, 1987), and streptomyces bacteriophages such as  $\phi$ C31 (Chater *et al.*, In: Sixth International Symposium on Actinomycetales Biology, Akademiai Kiado, Budapest, Hungary, pp. 45-54, 1986). *Pseudomonas* plasmids are reviewed by John *et al.* (*Rev. Infect. Dis.* 8:693-704, 1986), and Izaki (*Jpn. J. Bacteriol.* 33:729-742, 1978).

Preferred eukaryotic plasmids include, for example, BPV, vaccinia, SV40, 2-micron circle, and the like, or their derivatives. Such plasmids are well known in the art (Botstein *et al.*, Miami Wntr. Symp. 19:265-274, 1982; Broach, In: "The Molecular Biology of the Yeast *Saccharomyces*: Life Cycle and Inheritance", Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 445-470, 1981; Broach, *Cell* 28:203-204, 1982; Bollon *et al.*, *J. Clin. Hematol. Oncol.* 10:39-48, 1980; Maniatis, In: Cell Biology: A Comprehensive Treatise, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563-608, 1980).

Once the vector or nucleic acid molecule containing the construct(s) has been prepared for expression, the DNA construct(s) may be introduced into an appropriate host cell by any of a variety of suitable means, *i.e.*, transformation, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate-precipitation, direct microinjection, and the like. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene(s) results in the production of a kinase of the invention, or fragments thereof. This can take place in the transformed cells as such, or following the induction of these cells to differentiate (for example, by administration of bromodeoxyuracil to neuroblastoma cells or the like). A variety of incubation conditions can be used to form the peptide of the present invention. The most preferred conditions are those which mimic physiological conditions.

**Transgenic Animals:**

5 A variety of methods are available for the production of transgenic animals associated with this invention. DNA can be injected into the pronucleus of a fertilized egg before fusion of the male and female pronuclei, or injected into the nucleus of an embryonic cell (*e.g.*, the nucleus of a two-cell embryo) following the initiation of cell division (Brinster *et al.*, *Proc. Nat. Acad. Sci. USA* 82:4438-4442, 1985). Embryos can be infected with viruses, especially retroviruses, modified to carry inorganic-ion receptor nucleotide sequences of the invention.

10 Pluripotent stem cells derived from the inner cell mass of the embryo and stabilized in culture can be manipulated in culture to incorporate nucleotide sequences of the invention. A transgenic animal can be produced from such cells through implantation into a blastocyst that is implanted into a foster mother and allowed to come to term. Animals suitable for transgenic experiments can be obtained from standard commercial sources such as Charles River  
15 (Wilmington, MA), Taconic (Germantown, NY), Harlan Sprague Dawley (Indianapolis, IN), etc.

The procedures for manipulation of the rodent embryo and for microinjection of DNA into the pronucleus of the zygote are well known to those of ordinary skill in the art (Hogan *et al.*, *supra*). Microinjection procedures for fish, amphibian eggs and birds are detailed in Houdebine and Chourrout (*Experientia* 47:897-905, 1991). Other procedures for introduction of  
20 DNA into tissues of animals are described in U.S. Patent No. 4,945,050 (Sanford *et al.*, July 30, 1990).

By way of example only, to prepare a transgenic mouse, female mice are induced to superovulate. Females are placed with males, and the mated females are sacrificed by CO<sub>2</sub> asphyxiation or cervical dislocation and embryos are recovered from excised oviducts.  
25 Surrounding cumulus cells are removed. Pronuclear embryos are then washed and stored until the time of injection. Randomly cycling adult female mice are paired with vasectomized males. Recipient females are mated at the same time as donor females. Embryos then are transferred surgically. The procedure for generating transgenic rats is similar to that of mice (Hammer *et al.*, *Cell* 63:1099-1112, 1990).

30 Methods for the culturing of embryonic stem (ES) cells and the subsequent production of transgenic animals by the introduction of DNA into ES cells using methods such as

electroporation, calcium phosphate/DNA precipitation and direct injection also are well known to those of ordinary skill in the art (Teratocarcinomas and Embryonic Stem Cells, A Practical Approach, E.J. Robertson, ed., IRL Press, 1987).

5 In cases involving random gene integration, a clone containing the sequence(s) of the invention is co-transfected with a gene encoding resistance. Alternatively, the gene encoding neomycin resistance is physically linked to the sequence(s) of the invention. Transfection and isolation of desired clones are carried out by any one of several methods well known to those of ordinary skill in the art (E.J. Robertson, *supra*).

10 DNA molecules introduced into ES cells can also be integrated into the chromosome through the process of homologous recombination (Capecchi, *Science* 244:1288-1292, 1989). Methods for positive selection of the recombination event (*i.e.*, neo resistance) and dual positive-negative selection (*i.e.*, neo resistance and gancyclovir resistance) and the subsequent identification of the desired clones by PCR have been described by Capecchi, *supra* and Joyner *et al.* (*Nature* 338:153-156, 1989), the teachings of which are incorporated herein in their  
15 entirety including any drawings. The final phase of the procedure is to inject targeted ES cells into blastocysts and to transfer the blastocysts into pseudopregnant females. The resulting chimeric animals are bred and the offspring are analyzed by Southern blotting to identify individuals that carry the transgene. Procedures for the production of non-rodent mammals and other animals have been discussed by others (Houdebine and Chourrout, *supra*; Pursel *et al.*,  
20 *Science* 244:1281-1288, 1989; and Simms *et al.*, *Bio/Technology* 6:179-183, 1988).

Thus, the invention provides transgenic, nonhuman mammals containing a transgene encoding a kinase of the invention or a gene affecting the expression of the kinase. Such transgenic nonhuman mammals are particularly useful as an *in vivo* test system for studying the effects of introduction of a kinase, or regulating the expression of a kinase (*i.e.*, through the  
25 introduction of additional genes, antisense nucleic acids, or ribozymes).

A "transgenic animal" is an animal having cells that contain DNA which has been artificially inserted into a cell, which DNA becomes part of the genome of the animal which develops from that cell. Preferred transgenic animals are primates, mice, rats, cows, pigs, horses, goats, sheep, dogs and cats. The transgenic DNA may encode human kinases. Native

expression in an animal may be reduced by providing an amount of antisense RNA or DNA effective to reduce expression of the receptor.

### Gene Therapy:

5

Kinases or their genetic sequences will also be useful in gene therapy (reviewed in Miller, *Nature* 357:455-460, 1992). Miller states that advances have resulted in practical approaches to human gene therapy that have demonstrated positive initial results. The basic science of gene therapy is described in Mulligan (*Science* 260:926-931, 1993).

10

In one preferred embodiment, an expression vector containing a kinase coding sequence is inserted into cells, the cells are grown *in vitro* and then infused in large numbers into patients. In another preferred embodiment, a DNA segment containing a promoter of choice (for example a strong promoter) is transferred into cells containing an endogenous gene encoding kinases of the invention in such a manner that the promoter segment enhances expression of the endogenous kinase gene (for example, the promoter segment is transferred to the cell such that it becomes directly linked to the endogenous kinase gene).

15

The gene therapy may involve the use of an adenovirus containing kinase cDNA targeted to a tumor, systemic kinase increase by implantation of engineered cells, injection with kinase-encoding virus, or injection of naked kinase DNA into appropriate tissues.

20

Target cell populations may be modified by introducing altered forms of one or more components of the protein complexes in order to modulate the activity of such complexes. For example, by reducing or inhibiting a complex component activity within target cells, an abnormal signal transduction event(s) leading to a condition may be decreased, inhibited, or reversed. Deletion or missense mutants of a component, that retain the ability to interact with other components of the protein complexes but cannot function in signal transduction, may be used to inhibit an abnormal, deleterious signal transduction event.

25

Expression vectors derived from viruses such as retroviruses, vaccinia virus, adenovirus, adeno-associated virus, herpes viruses, several RNA viruses, or bovine papilloma virus, may be used for delivery of nucleotide sequences (*e.g.*, cDNA) encoding recombinant kinase of the invention protein into the targeted cell population (*e.g.*, tumor cells). Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors containing

30

coding sequences (Maniatis *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y., 1989; Ausubel *et al.*, Current Proto-cols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y., 1989). Alter-natively, recombinant nucleic acid mole-cules encoding protein sequences can be used as naked DNA or in a recon-stituted  
5 system *e.g.*, lipo-somes or other lipid systems for delivery to target cells (*e.g.*, Felgner *et al.*, *Nature* 337:387-8, 1989). Several other methods for the direct transfer of plasmid DNA into cells exist for use in human gene therapy and involve targeting the DNA to receptors on cells by complexing the plasmid DNA to proteins (Miller, *supra*).

In its simplest form, gene transfer can be performed by simply injecting minute amounts  
10 of DNA into the nucleus of a cell, through a process of microinjection (Capecchi, *Cell* 22:479-88, 1980). Once recombinant genes are introduced into a cell, they can be recognized by the cell's normal mechanisms for transcription and translation, and a gene product will be expressed. Other methods have also been attempted for introducing DNA into larger numbers of cells. These methods include: transfection, wherein DNA is precipitated with calcium phosphate and  
15 taken into cells by pinocytosis (Chen *et al.*, *Mol. Cell Biol.* 7:2745-52, 1987); electroporation, wherein cells are exposed to large voltage pulses to introduce holes into the membrane (Chu *et al.*, *Nucleic Acids Res.* 15:1311-26, 1987); lipofection/liposome fusion, wherein DNA is packaged into lipophilic vesicles which fuse with a target cell (Felgner *et al.*, *Proc. Natl. Acad. Sci. USA.* 84:7413-7417, 1987); and particle bombardment using DNA bound to small projectiles  
20 (Yang *et al.*, *Proc. Natl. Acad. Sci.* 87:9568-9572, 1990). Another method for introducing DNA into cells is to couple the DNA to chemically modified proteins.

It has also been shown that adenovirus proteins are capable of destabilizing endosomes and enhancing the uptake of DNA into cells. The admixture of adenovirus to solutions containing DNA complexes, or the binding of DNA to polylysine covalently attached to  
25 adenovirus using protein crosslinking agents substantially improves the uptake and expression of the recombinant gene (Curiel *et al.*, *Am. J. Respir. Cell. Mol. Biol.*, 6:247-52, 1992).

As used herein "gene transfer" means the process of introducing a foreign nucleic acid molecule into a cell. Gene transfer is commonly performed to enable the expres-sion of a particular product encoded by the gene. The product may include a protein, polypeptide, anti-  
30 sense DNA or RNA, or enzymatically active RNA. Gene transfer can be performed in cultured

cells or by direct administration into animals. Generally gene transfer involves the process of nucleic acid contact with a target cell by non-specific or receptor mediated interactions, uptake of nucleic acid into the cell through the membrane or by endocytosis, and release of nucleic acid into the cyto-plasm from the plasma membrane or endosome. Expression may require, in addition, movement of the nucleic acid into the nucleus of the cell and binding to appropriate nuclear factors for transcription.

As used herein "gene therapy" is a form of gene transfer and is included within the definition of gene transfer as used herein and specifically refers to gene transfer to express a therapeutic product from a cell *in vivo* or *in vitro*. Gene transfer can be performed *ex vivo* on cells which are then transplanted into a patient, or can be performed by direct administration of the nucleic acid or nucleic acid-protein complex into the patient.

In another preferred embodiment, a vector having nucleic acid sequences encoding a kinase polypeptide is provided in which the nucleic acid sequence is expressed only in specific tissue. Methods of achieving tissue-specific gene expression are set forth in International Publication No. WO 93/09236, filed November 3, 1992 and published May 13, 1993.

In all of the preceding vectors set forth above, a further aspect of the invention is that the nucleic acid sequence contained in the vector may include additions, deletions or modifications to some or all of the sequence of the nucleic acid, as defined above.

In another preferred embodiment, a method of gene replacement is set forth. "Gene replacement" as used herein means supplying a nucleic acid sequence which is capable of being expressed *in vivo* in an animal and thereby providing or augmenting the function of an endogenous gene which is missing or defective in the animal.

## 25 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

The compounds described herein can be administered to a human patient per se, or in pharmaceutical compositions where it is mixed with other active ingredients, as in combination therapy, or suitable carriers or excipient(s). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition.

**Routes Of Administration:**

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or  
5 intestinal administration; parenteral delivery, including intramuscular, subcutaneous,  
intravenous, intramedullary injections, as well as intrathecal, direct intraventricular,  
intraperitoneal, intranasal, or intraocular injections.

Alternately, one may administer the compound in a local rather than systemic manner, for  
example, via injection of the compound directly into a solid tumor, often in a depot or sustained  
10 release formulation.

Furthermore, one may administer the drug in a targeted drug delivery system, for  
example, in a liposome coated with tumor-specific antibody. The liposomes will be targeted to  
and taken up selectively by the tumor.

**15 Composition/Formulation:**

The pharmaceutical compositions of the present invention may be manufactured in a  
manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating,  
dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

Pharmaceutical compositions for use in accordance with the present invention thus may  
20 be formulated in conventional manner using one or more physiologically acceptable carriers  
comprising excipients and auxiliaries which facilitate processing of the active compounds into  
preparations which can be used pharmaceutically. Proper formulation is dependent upon the  
route of administration chosen.

For injection, the agents of the invention may be formulated in aqueous solutions,  
25 preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or  
physiological saline buffer. For transmucosal administration, penetrants appropriate to the  
barrier to be permeated are used in the formulation. Such penetrants are generally known in the  
art.

For oral administration, the compounds can be formulated readily by combining the  
30 active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers  
enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules,



liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Suitable carriers include excipients such as, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose,  
5 hydroxypropylmethyl- cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone,  
10 carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of  
15 gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition,  
20 stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration.

For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present  
25 invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of *e.g.* gelatin for use in

an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a cosolvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and

an aqueous phase. The cosolvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:D5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, *e.g.* polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose.

Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols.

Many of the tyrosine or serine/threonine kinase modulating compounds of the invention may be provided as salts with pharmaceutically compatible counterions. Pharmaceutically compatible salts may be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms.

**Suitable Dosage Regimens:**

Pharmaceutical compositions suitable for use in the present invention include compositions where the active ingredients are contained in an amount effective to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount of compound effective to prevent, alleviate or ameliorate symptoms of disease or prolong the survival of the subject being treated. Determination of a therapeutically effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein.

Methods of determining the dosages of compounds to be administered to a patient and modes of administering compounds to an organism are disclosed in U.S. Application Serial No. 08/702,282, filed August 23, 1996 and International patent publication number WO 96/22976, published August 1 1996, both of which are incorporated herein by reference in their entirety, including any drawings, figures or tables. Those skilled in the art will appreciate that such descriptions are applicable to the present invention and can be easily adapted to it.

The proper dosage depends on various factors such as the type of disease being treated, the particular composition being used and the size and physiological condition of the patient. Therapeutically effective doses for the compounds described herein can be estimated initially from cell culture and animal models. For example, a dose can be formulated in animal models to achieve a circulating concentration range that initially takes into account the  $IC_{50}$  as determined in cell culture assays. The animal model data can be used to more accurately determine useful doses in humans.

For any compound used in the methods of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the  $IC_{50}$  as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the tyrosine or serine/threonine kinase activity). Such information can be used to more accurately determine useful doses in humans.

Toxicity and therapeutic efficacy of the compounds described herein can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for

determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD<sub>50</sub> and ED<sub>50</sub>. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these  
5 cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician  
10 in view of the patient's condition. (See *e.g.*, Fingl *et al.*, 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1).

In another example, toxicity studies can be carried out by measuring the blood cell composition. For example, toxicity studies can be carried out in a suitable animal model as follows: 1) the compound is administered to mice (an untreated control mouse should also be  
15 used); 2) blood samples are periodically obtained via the tail vein from one mouse in each treatment group; and 3) the samples are analyzed for red and white blood cell counts, blood cell composition and the percent of lymphocytes versus polymorphonuclear cells. A comparison of results for each dosing regime with the controls indicates if toxicity is present.

At the termination of each toxicity study, further studies can be carried out by sacrificing  
20 the animals (preferably, in accordance with the American Veterinary Medical Association guidelines Report of the American Veterinary Medical Assoc. Panel on Euthanasia:229-249, 1993). Representative animals from each treatment group can then be examined by gross necropsy for immediate evidence of metastasis, unusual illness or toxicity. Gross abnormalities in tissue are noted and tissues are examined histologically. Compounds causing a reduction in  
25 body weight or blood components are less preferred, as are compounds having an adverse effect on major organs. In general, the greater the adverse effect the less preferred the compound.

For the treatment of cancers the expected daily dose of a hydrophobic pharmaceutical agent is between 1 to 500 mg/day, preferably 1 to 250 mg/day, and most preferably 1 to 50 mg/day. Drugs can be delivered less frequently provided plasma levels of the active moiety are  
30 sufficient to maintain therapeutic effectiveness.

Plasma levels should reflect the potency of the drug. Generally, the more potent the compound the lower the plasma levels necessary to achieve efficacy.

Plasma half-life and biodistribution of the drug and metabolites in the plasma, tumors and major organs can also be determined to facilitate the selection of drugs most appropriate to inhibit a disorder. Such measurements can be carried out. For example, HPLC analysis can be performed on the plasma of animals treated with the drug and the location of radiolabeled compounds can be determined using detection methods such as X-ray, CAT scan and MRI. Compounds that show potent inhibitory activity in the screening assays, but have poor pharmacokinetic characteristics, can be optimized by altering the chemical structure and retesting. In this regard, compounds displaying good pharmacokinetic characteristics can be used as a model.

Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the kinase modulating effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data; *e.g.*, the concentration necessary to achieve 50-90% inhibition of the kinase using the assays described herein. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%.

In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

#### **Packaging:**

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may

be accompanied by instructions for administration. The pack or dispenser may also be accompanied with a notice associated with the container in form prescribed by a governmental agency regulating the manufacture, use, or sale of pharmaceuticals, which notice is reflective of approval by the agency of the form of the polynucleotide for human or veterinary administration.

5 Such notice, for example, may be the labeling approved by the U.S. Food and Drug Administration for prescription drugs, or the approved product insert. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition. Suitable conditions indicated on the label may include treatment of a tumor, inhibition of

10 angiogenesis, treatment of fibrosis, diabetes, and the like.

#### *FUNCTIONAL DERIVATIVES*

15 Also provided herein are functional derivatives of a polypeptide or nucleic acid of the invention. By "functional derivative" is meant a "chemical derivative," "fragment," or "variant," of the polypeptide or nucleic acid of the invention, which terms are defined below. A functional derivative retains at least a portion of the function of the protein, for example reactivity with an antibody specific for the protein, enzymatic activity or binding activity mediated through

20 noncatalytic domains, which permits its utility in accordance with the present invention. It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. Equally, it is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or polypeptide that retains the functionality of the original. In both cases, all permutations are intended to be

25 covered by this disclosure.

Included within the scope of this invention are the functional equivalents of the herein-described isolated nucleic acid molecules. The degeneracy of the genetic code permits substitution of certain codons by other codons that specify the same amino acid and hence would give rise to the same protein. The nucleic acid sequence can vary substantially since, with the

30 exception of methionine and tryptophan, the known amino acids can be coded for by more than one codon. Thus, portions or all of the genes of the invention could be synthesized to give a

nucleic acid sequence significantly different from one selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57. The encoded amino acid sequence thereof would, however, be preserved.

In addition, the nucleic acid sequence may comprise a nucleotide sequence which results from the addition, deletion or substitution of at least one nucleotide to the 5'-end and/or the 3'-end of the nucleic acid formula selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57, or a derivative thereof.

Any nucleotide or polynucleotide may be used in this regard, provided that its addition, deletion or substitution does not alter the amino acid sequence of selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57, or a derivative thereof.



NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, and SEQ ID NO:57 which is encoded by the nucleotide sequence. For example, the present invention is intended to include any nucleic acid sequence resulting from the addition of ATG as an initiation codon at the 5'-end of the inventive nucleic acid sequence or its derivative, or from the addition of TTA, TAG or TGA as a termination codon at the 3'-end of the inventive nucleotide sequence or its derivative. Moreover, the nucleic acid molecule of the present invention may, as necessary, have restriction endonuclease recognition sites added to its 5'-end and/or 3'-end.

Such functional alterations of a given nucleic acid sequence afford an opportunity to promote secretion and/or processing of heterologous proteins encoded by foreign nucleic acid sequences fused thereto. All variations of the nucleotide sequence of the kinase genes of the invention and fragments thereof permitted by the genetic code are, therefore, included in this invention.

Further, it is possible to delete codons or to substitute one or more codons with codons other than degenerate codons to produce a structurally modified polypeptide, but one which has substantially the same utility or activity as the polypeptide produced by the unmodified nucleic acid molecule. As recognized in the art, the two polypeptides are functionally equivalent, as are the two nucleic acid molecules that give rise to their production, even though the differences between the nucleic acid molecules are not related to the degeneracy of the genetic code.

A "chemical derivative" of the complex contains additional chemical moieties not normally a part of the protein. Covalent modifications of the protein or peptides are included within the scope of this invention. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues, as described below.

Cysteinyl residues most commonly are reacted with alpha-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing primary amine containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high  $pK_a$  of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine alpha-amino group.

Tyrosyl residues are well-known targets of modification for introduction of spectral labels by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimide ( $R'-N-C-N-R'$ ) such as 1-cyclohexyl-3-(2-morpholinyl(4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Glutaminyl and asparaginy residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Derivatization with bifunctional agents is useful, for example, for cross-linking the  
5 component peptides of the protein to each other or to other proteins in a complex to a water-insoluble support matrix or to other macromolecular carriers. Commonly used cross-linking agents include, for example, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and  
10 bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl) dithiolpropioimide yield photoactivatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are  
15 employed for protein immobilization.

Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (Creighton, T.E., *Proteins: Structure and Molecular Properties*, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine,  
20 and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the stability, solubility, absorption, biological half life, and the like. The moieties may alternatively eliminate or attenuate any undesirable side effect of the protein complex and the like. Moieties capable of mediating such effects are disclosed, for example, in Remington's *Pharmaceutical Sciences*, 18th ed., Mack Publishing Co.,  
25 Easton, PA (1990).

The term "fragment" is used to indicate a polypeptide derived from the amino acid sequence of the proteins, of the complexes having a length less than the full-length polypeptide from which it has been derived. Such a fragment may, for example, be produced by proteolytic cleavage of the full-length protein. Preferably, the fragment is obtained recombinantly by  
30 appropriately modifying the DNA sequence encoding the proteins to delete one or more amino

acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence. Fragments of a protein are useful for screening for substances that act to modulate signal transduction, as described herein. It is understood that such fragments may retain one or more characterizing portions of the native complex. Examples of such retained characteristics include:

5 catalytic activity; substrate specificity; interaction with other molecules in the intact cell; regulatory functions; or binding with an antibody specific for the native complex, or an epitope thereof.

Another functional derivative intended to be within the scope of the present invention is a "variant" polypeptide which either lacks one or more amino acids or contains additional or substituted amino acids relative to the native polypeptide. The variant may be derived from a

10 naturally occurring complex component by appropriately modifying the protein DNA coding sequence to add, remove, and/or to modify codons for one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence. It is understood that such variants having added, substituted and/or additional amino acids retain one or more

15 characterizing portions of the native protein, as described above.

A functional derivative of a protein with deleted, inserted and/or substituted amino acid residues may be prepared using standard techniques well-known to those of ordinary skill in the art. For example, the modified components of the functional derivatives may be produced using site-directed mutagenesis techniques (as exemplified by Adelman *et al.*, 1983, *DNA* 2:183)

20 wherein nucleotides in the DNA coding the sequence are modified such that a modified coding sequence is modified, and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host cell, using techniques such as those described above. Alternatively, proteins with amino acid deletions, insertions and/or substitutions may be conveniently prepared by direct chemical synthesis, using methods well-known in the art. The functional derivatives of the

25 proteins typically exhibit the same qualitative biological activity as the native proteins.

TABLES  
AND  
DESCRIPTION THEREOF

5           Table 1 documents the name of each gene, the classification of each gene, the positions of the open reading frames within the sequence, and the length of the corresponding peptide. From left to right the data presented is as follows: "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family", "NA\_length", "ORF Start", "ORF End", "ORF Length", and "AA\_length". "Gene name" refers to name given the sequence encoding the kinase or kinase-  
10 like enzyme. Each gene is represented by "SGK" designation followed by a number. The SGK name usually represents multiple overlapping sequences built into a single contiguous sequence (a "contig"). The "ID#na" and "ID#aa" refer to the identification numbers given each nucleic acid and amino acid sequence in this patent. "FL/Cat" refers to the length of the gene, with FL indicating full length, and "Cat" indicating that only the catalytic domain is presented. "Partial"  
15 in this column indicates that the sequence encodes a partial protein kinase catalytic domain. "Superfamily" identifies whether the gene is a protein kinase or protein-kinase-like. "Group" and "Family" refer to the protein kinase classification defined by sequence homology and based on previously established phylogenetic analysis [Hardie, G. and Hanks S. The Protein Kinase Book, Academic Press (1995) and Hunter T. and Plowman, G. Trends in Biochemical Sciences  
20 (1977) 22:18-22 and Plowman G.D. et al. (1999) Proc. Natl. Acad. Sci. 96:13603-13610)]. "NA\_length" refers to the length in nucleotides of the corresponding nucleic acid sequence. "ORF start" refers to the beginning nucleotide of the open reading frame. "ORF end" refers to the last nucleotide of the open reading frame, excluding the stop codon. "ORF length" refers to the length in nucleotides of the open reading frame (excluding the stop codon). "AA length"  
25 refers to the length in amino acids of the peptide encoded in the corresponding nucleic acid sequence.

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	NA-length	ORF Start	ORF End	ORF Length	AA-length	Genomic ID	sort1	Ortholog	Header	Repeats	SNPs
SGK187_CRIK	1	58	FL	PK	AGC	DMPK	6159	1	8159	8159	2053	4926912 1 1	44	AAC72823.1: AAC25483.1:18100 0000994928:AC002 583.1	>CRIK-	none	4035=Y (ccgagagagcgc agagagagagcgc ); 3327=R (gggagagagagc gagagagagcgc ); 965=K (cagagagagagc agagagagagcgc ); 1318=R (gagagagagagc agagagagagcgc ); 316=K
SGK084_GRK7	2	59	FL	PK	AGC	GRK	1682	1	1659	1659	553	11000284009825	23	AAC9500	>GRK7 -	285 ggagagagagagc gagagagagagc gagagagagagc	1318=R (gagagagagagc agagagagagagc ); 316=K
SGK409_KIAA0303	3	60	FL	PK	AGC	MAST	7785	1	7569	7569	2523	6871897 1 1	54	NP_032667 BAA76817	>SGK409 SEQ_ID_0 2_687189 7_11: FLV predicted from contig of SGK409 (at 5' end) and KIAA0303 (at 3' end)	2355 ggagagagagagc gagagagagagc gagagagagagc	8327=M (ccgagagagagc agagagagagagc ); 6282=M (cagagagagagc gagagagagagc ); 2375=M
SGK021	4	61	Cat	PK	AGC	Mo3C11.1 ce	981	1	981	981	327	17000028181153	8	CAB76566 BAA60277	>SGK021 SGK225_1 Incy_1110 937.1: Genewise prediction: model: STK_h NP_056877 1: target: 1730001 5832053	9-S (gagagagagc; ); 974=K (gagagagagagc ); 2375=M	

Table 1 - Open Reading Frames  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	NA length	ORF Start	ORF End	ORF Length	AA length	Genomic ID	sort1	Ortholog	Header	Repeats	SNPs
SGK410	5	62	FL	PK	AGC	PKC	2260	72	1964	1893	631	7018620 2	55	NP 002731	>SGK410 SEQ_ID_0 3_682189 8_50001_ 8; FL from Genewise prediction from HGP contig AL133280 .12 using PKC.jala_ h NP_002731 1.1 as a model	none 1656 GAGAGGAGAG GGGAGAGAGAG a 1678	
SGK069	8	63	Cat	PK	AGC	Unique	1716	1	1716	1716	572	11000283338966	28	BAA36362	>SGK069		

Gene Name	ID#na	ID#waa	FL/Cat	Super-family	Group	Family	NA_length	ORF Start	ORF End	ORF Length	AA_length	Genomic ID	sort1	Ortholog	Header	Repeats	SHPs
SGK110	7	64	partial	PK	AGC	Unique	1118	1	1119	1118	373	11000284702500	35	BAA36362	>SGK110 1100283 338966 and 17000077 607693	none	
SGK053_CKLJK	8	65	FL	PK	CAMK	AMPK	1074	1	1071	1071	357	17000035780284	20	Q63450	>SGK053 13-AUG- 2000 NM_0203 97.1 - Casein kinase-like 419 klnsqcagqgql alscgaactggg ccacagaacc 440 angfcaam)	605=Y (aaccttcacgcga agcg Y); 509=M lgnicagcagga ccacagaacc 440 angfcaam)	
SGK124	9	66	FL	PK	CAMK	AMPK	1077	1	1074	1074	358	17000047891899	37	p.Lr	>SGK124 SEQ_ID_0 4_554186 1_50001_ 6 G1726390 4lembIALD 34548.25] HS1103G 7.[142689; 151365].s	188=S (tccagataglcga adcgagaggSt) 333=Y (tgpcaglgcdga gccaaaY)	
SGK254_CAMKKa	10	67	FL	PK	CAMK	CAMK	1542	1	1539	1539	513	4454511_1_3	47	AAB46910	>SGK254 SEQ_ID_0 7_445451 1_1_3 G11445451 11gbiACOO 5940.3IAC 005940.[9 4560:118D 33]sp	1148=R (lpgclttccddagp R)	
SGK297_CAMKIb2	11	68	FL	PK	CAMK	CAMK	1032	1	1029	1029	343	17000113122640	50	JL	>CAMKIb2 JL INCX_BZT 43ICB1--	77=K (lppdglgdagap torpk)	



Table 1 - Open Reading Frames  
413406\_1.xls

Gene Name	IDna	IDaa	FL/Cat	Super-family	Group	Family	NA length	ORF Start	ORF End	ORF Length	AA length	Genomic ID	sort1	Ortholog	Header	Repeats	SNPs
SGK411, CamKII delta2	12	69	FL	PK	CAMK	CAMK	1500	1	1497	1497	499	2828765 1 3	56	AA020442	>SGK411 SEQ_ID_0 6_282876 5_1_3: FL from CamKII delta2_h AF071569 18-JAN- 2000—	none	15-H: agggcagaccac M
SGK027	13	70	FL	PK	CAMK	EMK	1311	1	1308	1308	436	17000029868654	8	AAF69801 AA07437	>SGK027 CA2 GS N_106000 01135116 7_1 17000057 443791 on N-term	none	45-Y lgagagagcagp cdaglgagaccac 2Y
SGK046b	14	71	partial	PK	CAMK	EMK	225	1	225	225	75	11000284253087	16	AAC33487 AAC33487 AAF69801	>SGK046 b	none	
SGK046c	15	72	partial	PK	CAMK	EMK	117	1	117	117	39	11000284253087	17	AAC33487 AAC33487 AAF69801	>SGK046c	none	
SGK089	16	73	partial	PK	CAMK	EMK	252	1	252	252	84	11000283986588	32	AAF64455 AAA97437	>SGK089 13-AUG- 2000 AK024110 (dimer)?	none	2003-S (gdcacagagcagp gacacagagcagp GS), 1673-S (cacacacacac cagagagS)
SGK133	17	74	FL	PK	CAMK	EMK	2385	1	2382	2382	794	6671694 1 1	40	CAA07198	>SGK133 6910558_2 INCY_984 011.1 -	none	
SGK004, MSK	18	75	FL	PK	CAMK	EMK-1	2381	1	2359	2358	786	1100025774647	3	>SGK004	>SGK004	none	
SGK006	19	76	Cat	PK	CAMK	EMK-1	789	1	789	789	263	11000257829454	4	NP 056570	>SGK006 13-AUG- 2000 = SGK006.8 gw-falid= 10353)_P RED_CDIN A_GW_H OMO_len (ghr789-	none	

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	NA_length	ORF Start	ORF End	ORF Length	AA_length	Genomic ID	sort1	Ortholog	Header	Repeats	SNPs
SGK180, SNRK	20	77	FL	PK	CAMK	EMK-1	2298	1	2295	2295	765	1700003577785	43			1332 ggggagagaga lgaagagaa 1353	1817+5 (ggccccggag arccaaagcggg gg)
SGK386, MLCKs	21	78	FL	PK	CAMK	MLCK	1836	1	1836	1836	612	17000140438285	51			>SNRK_h- h SGK386 128515.1 gi 821764 6emb AL1 60175.5 A L160175.1 114387.12 65801.sp HGP_724 2443_3 plus Celera gi 821764 6emb AL1 60175.5 A L160175.1 114387.12 65801.sp.17 none	835-M (tttggagatgac ggccaa)
SGK003	22	78	FL	PK	CKI	CKI	1014	1	1014	1014	337	11000257741254	2	P70065		none 488	
SGK066	23	80	Cal	PK	CKI	CKI	1200	1	1200	1200	400	11000283296349	24	T42260		>SGK066 lgaatggagagc tga 504	

Table 1 - Open Reading Frames  
413406\_1.xls

Gene Name	IDna	IDaa	FL/Cat	Super-family	Group	Family	NA_length	ORF Start	ORF End	ORF Length	AA_length	Genomic ID	sort1	Ortholog	Header	Repeats	SNP's
SGK041, NKIAMRE	24	81	FL	PK	CMGC	CDK	1773	1	1773	1773	591	11000257953829	15		>NKIAMR E AF130372 11AF1303 72 Homo sapiens serine- threonine protein kinase (NKIAMR E) mRNA, complete cdfs	1002-R (caccagcagcagc aggagat; 1284-R (ccacatgcgga ggctgcgG); 1161-Y (gcaagagaaagc tccacaY)	
SGK112	25	82	FL	PK	CMGC	CDK	1083	1	1080	1080	360	17000035915087	36	S22745	>SGK112 14-AUG- 2000	none	
SGK038, ERK7	26	83	FL	PK	CMGC	MAPK	1677	1	1674	1674	558	17000030278391	13	AAD12719	>SGK038 13-AUG- 2000	508 cgaagcagcagc tccagc 528	
SGK158	27	84	FL	PK	Microbial PK	ABC1	2256	1	2253	2253	751	11000283551487	42	Q60936 Q92339	>SGK158 Incy40605 7-19-	1216 (accagagcagcagc aggagagacagc tccaca 1285 Y)	1751-Y (accagagcagcagc aggagagagcagc tccaca 1285 Y)
SGK429	28	85	FL	PK	Microbial PK	ABC1	1881	1	1878	1878	626	5001549_3	60	CAB76039	>SGK429 HGP, Incy05256 0.1-	1630 accagagcagcagc aggagagacagc tccaca 1285 1531	186-R (ccacagagcagcagc aggagagacagc tccaca 1285 Y)
SGK152, SUDD	29	86	FL	PK	Microbial PK	RID1	1560	1	1557	1557	519	11000258249295	41		>SGK152-	none	(AAACTAAAT CCACGTAAAG ATCA)
SGK077	30	87	FL	PK	Other	C28C2 ce	2397	1	2394	2394	798	11000283672012	29	NP_034483	>SGK077 1157803:1 601961.sp	877-K (accagagcagcagc aggagagacagc tccaca 1285 Y)	

Table 1 - Open Reading Frames  
413406\_1.xls

Gene Name	ID#s	ID#s	FL/Cat	Super-family	Group	Family	NA_length	ORF Start	ORF End	ORF Length	AA_length	Genomic ID	sort1	Ortholog	Header	Repeats	SNPs
SGK093, Wnk3	31	88	FL	PK	Other	C25C2 ce	4542	1	4539	4539	1513	11000284063441	33	AAF74258	>SGK093 14-AUG- 2000; FLV from Genescan prediction from Celera 96000001 680843; corrected manually by deletion of aa sequence *HRLGEC WQKMR RQGGAA GGNEPV GGSPPE DVSPHQ DSGYAPS PR* which did not fit the HMM prediction	none	2495-K (gccccaccccc gaagccak); 3295-M (caaaatatacagc aacatlm)
SGK074	32	89	Cal	PK	Other	DYRK	1065	1	1065	1065	355	11000283476699	27	NP 034562	>SGK074 13-AUG- 2000	none	
SGK087	33	90	FL	PK	Other	DYRK	2112	1	2109	2109	703	17000036171996	31	AAD47290 NP 003574	>SGK087 14-AUG- 2000	1284 caggaaccag gagagga 1283	
SGK295, KIS	34	91	FL	PK	Other	DYRK	1260	1	1257	1257	419	17000113079883	49	NP 058989	>SGK295- W	3355-W (caggaaccagc w)	
SGK419	35	92	FL	PK	Other	NAK	1866	1	1983	1983	661	5932598_100001_1	57	CAB70863 NP 055726	>SGK419 SEQ_ID 2 5_593259 8_100001 _1: FL from DKFZp43 4P016.h NM_0175 93_04- JUL 2000-	none	2447-K (atcaaaatctctd aaacacagaaaa ttc)

Table 1 - Open Reading Frames  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	NA length	ORF Start	ORF End	ORF Length	AA length	Genomic ID	sort1	Ortholog	Header	Repeats	SNPs
SGK125, MYO3A	36	93	FL	PK	Other	NineC	4848	1	4845	4845	1815	17000078080910	38		>MYO3A, NM_0174 33.1 Hom ocapiens myosinIIA mRNA-	none	3145+Y (attaccatctggaag capitaaatY); 3204+Y (gattcaacatgagc capitaaatY)
SGK445	37	94	partial	PK	Other	PLK	204	1	204	204	68	4000001803382	61	NP_055078/590000 28893040 (11.5 kb)	>SGK445-	183 aaas200	
SGK127	38	95	FL	PK	Other	RAF	2838	1	2835	2835	945	17000062804843	39	>NP_038599.1 kna se_suppressor_of_ as [Mus_musculus] B1 -	>SGK127 17000062 804843, lncv_1041 823, lncvle_11 03660, 17000140 021687, 7276286 lncvle 7474604C	none 427	
SGK009, ANKRD3	39	96	FL	PK	Other	RIP	2499	1	2496	2496	832	11000258161587	6		>SGK009	gaactcaagacc ggacacacagac 1452	
SGK421, STK22A, TSK1	40	97	FL	PK	Other	STK22A	1104	1	1101	1101	357	6759077 1 4	59	NP_033461	>SGK421 40000018 03623-	none	
SGK047	41	98	partial	PK	Other	STKR	93	1	93	93	31	11000256262374	18	NP_004320	>SGK047	none	89-R (cggagatctgga cdgaglaaacta R)
SGK196	42	99	FL	PK	Other	Unique	1053	1	1050	1050	350	17000062537825	46	T01289	>SGK196	none	
SGK396	43	100	FL	PK	Other	Unique	1419	1	1416	1416	472	5630059 1 4	53	CAB90410 CAA11152	>SGK396 SEQ_ID_3 1_563005 9 1 4	none	
SGK279, PKN	44	101	FL	PK	Other	YWY3 ca	1275	1	1272	1272	424	17000097259742	48	BAA36362 S71887	>PKN-	none	

Table 1 - Open Reading Frames  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	NA length	ORF Start	ORF End	ORF Length	AA length	Genomic ID	sort1	Ortholog	Header	Repeats	SNPs
SGK037	45	102	FL	PK	STE	NEK	1947	1	1947	1947	649	17000030265658	12	P51856, P51954	>SGK037, SGK113, SGK039 13-AUG- 2000; FL virtual from genewise/ genescan of celera assembly 17300003 627483BM DK start from HGP 6996171 (positions 144536- 114544) - still missing HPN motif Cterm predicted from genescan of celera	none 145	
SGK080	46	103	FL	PK	STE	NEK	1938	1	1935	1935	645	17000036897142	22	NP 035979	>SGK080 747758SC B1: 13- AUG-2000	1463-W 1465 1467 (gaugcagc cangggw)	
SGK080	47	104	FL	PK	STE	NEK	1338	1	1338	1338	446	17000048119006	30	NP 002488	>SGK080 13-AUG- 2000	none	

Table 1 - Open Reading Frames  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	NA_length	ORF Start	ORF End	ORF Length	AA_length	Genomic ID	sort1	Ortholog	Header	Repeats	SNPs
SGK002	48	105	FL	PK	STE	STE11	1829	327	1523	1197	399	11000257702871	1	P36507	>SGK002 SEQ_ID_3 5.667197 3.1.4; FL from Genewise prediction from HGP AC018639 .8 using MPK2_h P36507 as a model; note that this gene is intronless and encodes a stop (position x) found in HGP and celera genomic seq as well as	none 733 gaaaagcag capactg 751	1165=R (gagagagacac ggccg) 1753=R (gggacagag agggacg) 883=M (GGGGTGAG CGGCCAGCT CATM)
SGK058	49	106	Cat	PK	STE	STE11	834	1	834	834	278	17000035877198	21	A48084 AAC97114	>SGK058 13-AUG- 2000 -	none	
SGK103	50	107	partial	PK	STE	STE11	84	1	84	84	28	11000284272557	34	CAA39285 AAA28552	>SGK103 14-AUG- 2000	none	
SGK035	51	108	partial	PK	STE	STE20	1044	1	1044	1044	348	17000030169905	11	Q13177	>SGK035 481144.6 genewise on C-term	133 cagaaagagaa gaaaga 151: 665 adgaagacac (cagmtacgga agaag 713 67)	273=Y (cagmtacgga agaag 713 67)
SGK075	52	109	Cat	PK	STE	STE20	3318	1	3318	3318	1106	11000283492249	28	NP_059129 P50527	>SGK075	none	877=R (mdgcaaaaac gagggcgt)
SGK188_Epha9	53	110	FL	PK	TK	RTK-11	3112	74	3100	3027	1009	1700005739161	45		>SGK188 Incyla_74 74721CB1	none	2319=R (gagagagagac gagggcgt) 1888=R (gagggcagag gagggcaggt) & 1004=Y (gagggcagag gagggcaggt) 480 mdgcaaaaac (gagggcaggt agaag 478 68c7)
SGK040	54	111	FL	PK	TK	Unique	2730	1	2727	2727	909	11000257912897	14	P41243 T33475	>SGK040 SEQ_ID_4 0.622704 10 1 1	480 mdgcaaaaac (gagggcaggt agaag 478 68c7)	

Gene Name	ID#na	FL/Cat	Super-family	Group	Family	NA_length	ORF Start	ORF End	ORF Length	AA_length	Genomic ID	sort1	Ortholog	Header	Repeats	SNPs
SGK390	55	FL	PK-like	DAG kin	DAG kin	3495	1	3492	3492	1164	17000112562523	52	Q64398	>SGK390 14-AUG-2000	none	1314+R (gagagagacata gagagagacata gagagagacata R)
SGK007	56	Cal	PK-like	GCyc	GCyc	2652	1	2652	2652	884	11000257855312	5	T42260	>SGK007	none	
SGK050	57	partial	PK-like	GCyc	GCyc	144	1	144	144	48	17000035767558	19	P16067	>SGK050 13-AUG-2000	none	



Table 2 lists the following features of the genes described in this application: chromosomal localization, single nucleotide polymorphisms (SNPs), representation in dbEST, and repeat regions. From left to right the data presented is as follows: "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family", "Chromosome", "SNPs", "dbEST\_hits", & "Repeats". The contents of the first 7 columns (i.e., "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family") are as described above for Table 1. "Chromosome" refers to the cytogenetic localization of the gene. Information in the "SNPs" column describes the nucleic acid position and degenerate nature of candidate single nucleotide polymorphisms (SNPs). For example, for SGK386, the "SNPs" column contains "835=M", indicating that there are instances of both a C and an A (M = C or A) at position 835. "dbESThits" lists accession numbers of entries in the public database of ESTs (dbEST, <http://www.ncbi.nlm.nih.gov/dbEST/index.html>) that contain at least 100 bp of 100% identity to the corresponding gene. These ESTs were identified by blastn of dbEST. "Repeats" contains information about the location of short sequences, approximately 20 bp in length, that are of low complexity and that are present in several distinct genes. These repeats were identified by blastn of the DNA sequence against the non-redundant nucleic acid database at NCBI (nrna). To be included in this repeat column, the sequence typically could have 100% identity over its length and typically is present in at least 5 different genes.

Table 2 - CHR, SNPs, dbEST, Repeats  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Chromosome	SNPs	dbEST_hits	Repeats	Genomic_ID	sort1	Ortholog	Header	Repeats	SNPs
SGK187, CRIK	1	58	FL	PK	AGC	DMPK	12q24.23	2874=R (s1337340); 2883=Y (r904855); 3327=R (s1581824)	BE00488, BE87587, AW60530	none	4926912_1_1	44	AAC72823.1 : AAC25483.1 :181000000 994928:ACO 02563.1	>CRIK-	none	3327=R (ggttcgagagcgcagagc) dbSNP s1581824; 2883=Y (agagcagcagcagagc) dbSNP r904855; 2874=R (gggagagagagcagc) dbSNP s1337340; 985=Y (cagagcgcgcagagc); 1318=R (agagcagcagagcagc R)
SGK064, GRK7	2	59	FL	PK	AGC	GRK	3q24	9654=K; 1318=R	none	285 - 314	11000284009826	23	AAC9500 >GRK7 -		285 ggagccagcagagcagc 314	
SGK409, KIAA0303	3	60	FL	PK	AGC	MAST	5q12.1	6282=M; 6327=M	BE515326, BE287204, AA826842	2355 - 2375	6671897_1_1	54	>SGK409 SEQ_ID_02 _5671897_1 _1: FLV predicted from contig of SGK409 (at 5' end) and KIAA0303 NP_032667 BAA76817 (at 3' end)-		2355 ggagagcagcagagcagc 2375	6327=M (cctccagagcagcagc ccaa); 6282=M (cagagcagcagcagc agcagc)
SGK021	4	61	Cat	PK	AGC	Mo3C11.1 _c8	5q31.2	9=S; 97=R	none	none	17000028181153	8	>SGK021 SGK225. locus_111003 Z1. Genewise prediction: model: STK_b NP_068671: target: CAB76566 BAA95027 33053-		none	9=S (gagagcagcagcagc) (agcagcagcagcagc)

Table 2 - CHR, SNPs, dbEST, Repeats  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Chromosome	SNPs	dbEST_hits	Repeats	Genomic_ID	sort1	Ortholog	Header	Repeats	SNPs
SGK410	5	62	FL	PK	AGC	PKC	Xq23		none	none	7018620_2	55	NP_002731	>SGK410 SEQ_ID_03 _6821956_5 0001_6; FL from Genewise prediction from HGP contig AL133280.1 PKC iota_h NP_002731. 1 as a model	none	1180=S (cagttgagcagccgcs) dbSNP ss1317628: 210=Y (tcacagagagagY) dbSNP ss186813
SGK069	6	63	Cat	PK	AGC	Unique	19p11.13	1180=S (ss1317628): 210=Y (ss1668813)	none	1658-1678	11000283338966	26	BAA36362	>SGK069- >SGK110 1100028333 8966 and 1700007760 7693	1656 gacacagagagagagagag aaa 1678	1180=S (cagttgagcagccgcs) dbSNP ss1317628: 210=Y (tcacagagagagY) dbSNP ss186813
SGK110	7	64	partial	PK	AGC	Unique	19q13.4	597=R (r664439): 252=Y (ss661406)	none	none	11000284702500	35	BAA36362	>SGK053 13-AUG- 2000 NM_020397. 1 - Casein kinase-like	none	597=R (aagtgagagagagagcc) dbSNP rs554438: 252=Y (cagagagagagY) dbSNP ss661406
SGK053, CKLIK	8	65	FL	PK	CAMK	AMPK	10p14	805=Y; 509=M (ss661406)	BE260895, A023704, AW501047	419 - 440	17000035790284	20	O63450	>SGK124 SEQ_ID_04 _5541861_5 0001_6 p17263904 embJAL0345 48.25[H511 Q3G7.1[426 89:151365]. sp.Y	419 tgcacacagagagagagag 440	805=Y (aactatagagagagagY); 509=M (aactatagagagagagY) (aactatagagagagagY)
SGK124	9	66	FL	PK	CAMK	AMPK	20p12.2-p13	188=S; 333=Y (ss661406)	BF028145, BE260893	none	17000047891899	37		>SGK254 SEQ_ID_07 _4454511_1 _3 g14454511 gbJAC00594 0.3JAC0059 40194560:1	none	188=S (tccagatagagagagagag S); 333=Y (tccagatagagagagagY)
SGK254, CAMKka	10	67	FL	PK	CAMK	CAMK	17p13.3	555=R (ss661406): 1148=R	BE261149	none	4454511_1_3	47	AAB46910	1803315P	none	555=R (tccagatagagagagY) dbSNP ss661406: 1148=R (tccagatagagagagY)

[illegible]

Table 2 - CHR, SNPs, dbEST, Repeats  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Chromosome	SNPs	dbEST_hits	Repeats	Genomic_ID	sort1	Ortholog	Header	Repeats	SNPs
SGK386, MLCKs	21	78	FL	PK	CAMK	MLCK	20q11.1	835=M	AA197072, R02824	none	17000140438265	51		>MLCKs_h SGK386 126515.1 gi 8217646  embJAL1601 75.5JAL1601 75.1114367: 128560 sp HGP_72424 43_3 plus Celera gi 8217646  embJAL1601 75.5JAL1601 75.1114367: 128560 sp.1 r	none	835=M (entrezgigantocodcaM)
														>SGK003 13-AUG- 2000: FL genewise prediction from celera sequence 7800000630 66170 using CKIalpha_p g P70065 as a model; manually corrected by addition of seq gatactaa at 1006-1014 based on pairwise comparison between aa seq for SGK003 and CKIalpha_h NP_001883. 2 and CKI,		
SGK003	22	79	FL	PK	CKI	CKI	13q14.11		none	none	11000257741254	2	P70065		none	
SGK066	23	80	Cat	PK	CKI	CKI	15q15		AA324451	488-504	11000283296349	24	T42260	>SGK066	488 1gattggcgggacga 504	

Table 2 - CHR, SNPs, dbEST, Repeats  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Chromosome	SNPs	dbEST_hits	Repeats	Genomic_ID	sort1	Ortholog	Header	Repeats	SNPs
SGK041, NKIAMRE	24	81	FL	PK	CMGC	CDK	5q31.1	1033=R; 1181=Y;1284=R	AI894625, AI894352, AI812547	none	11000257953829	15		>NKIAMRE AF130372.1 AF130372 Homo sapiens serine-threonine protein kinase (NKIAMRE) mRNA, complete cds	1033=R (cgaagagcagtttgggaR); 1284=R (tcacatctcggaggttcgR); 1181=Y (gcaatgaaaatgctatcda Y)	
	25	82	FL	PK	CMGC	CDK	CHR2		AA828859	none	17000035915087	36	S22745	>SGK112 14-AUG-2000	none	
SGK038, ERK7	26	83	FL	PK	CMGC	MAPK	na		BE464560, AI048667	508 - 528	17000030278391	13	AAD12719	>SGK038 13-AUG-2000	508 gtgaagcagcgaactgtgc 528	
SGK158	27	84	FL	PK	Microbial PK	ABC1	1q42.11-42.2	1752=Y (ss1529338)	BE707060, AY006871, AI819411	1246 - 1265	11000283551497	42	Q60936 Q92338	>SGK158 IncY406057_19-	1246 atgatgaaactctcaaaa 1265	1752=Y (atcctgtctctgtctcagggaa gctgttY) dbSNP ss1529338
SGK429	28	85	FL	PK	Microbial PK	ABC1	7q34-35	196=R; 919=Y (ss1549835); 1865=Y (ss1517749)	BE877541, BE745459, BE259124	1830 - 1851	5001549_3	60	CAB76039	>SGK429 HGP, IncY052560_1-	1830 acgagatagcctcagcagga 1851	196=R (ccccagctgrr); 919= (agggcacaactcttcY) dbSNP ss1549835; 1865=Y (CCTCCTCACAGGGCCY Y) dbSNP ss1517749; 972=M
SGK152, SUDD	29	86	FL	PK	Microbial PK	R101	18p11.1	972=M	BE821277, AA452706	none	11000258249295	41		>SGK152- AGATCM	none	(AACTAAATCCACGTA AGATCM) 390=Y (agtcaccccY) dbSNP ss1659885; 811=R (cagcgtccagagrr) dbSNP ss1629760; 985=Y (ttgtgcaggggaaY) dbSNP ss1629759
SGK077	30	87	FL	PK	Other	C26C2_c9	17p13.3	390=Y (ss1659885); 611=R (ss1629760); 985=Y (ss1629759)	AA504563, AW752337	none	11000283672012	29	NP_034483	>SGK077 [157803:160195] sn	none	

Table 2 - CHR, SNPs, dbEST, Repeats  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Chromosome	SNPs	dbEST_hits	Repeats	Genomic_ID	sort1	Ortholog	Header	Repeats	SNPs
SGK093, Wnk3	31	88	FL	PK	Other	C26C2.2	17q21.1-2	3279 = K 4078 = M	BF000389, AI60766, AA87258	none	11000284063441	33	AAF74258	>SGK093 14-AUG- 2000; FLV from Genescan Prediction from Celera 9600000168 0843: corrected manually by deletion of aa sequence *HRLGECW QKMRRLQ QGAAGGN FPVGGSEF EDVSPHDD SGYAPSPR * which did not fit the HMM prediction and did not align to NEK7.	none	3279 = K (gcccccaacccctgagccac) : 4078 = M (caaaatcatcagctcacatgta)
SGK074	32	89	Cat	PK	Other	DYRK	19p12-19q13		none	none	11000283476699	27	NP_034562	>SGK074 13-AUG- 2000	none	
SGK087	33	90	FL	PK	Other	DYRK	12p13.3	269 = R (ss88136) 355 = W	BE243965 BE65118, AI21234, BE145607	1264 - 1283	17000036171996	31	AA047280 NP_003574	>SGK087 14-AUG- 2000	1264 cagatcccaagagcagagga 1283	269 = R (acttccacagcagcagagc) dbSNP 158138
SGK295, KIS	34	91	FL	PK	Other	DYRK	1q23.3			none	17000113079883	49	NP_056989	>SGK295- >SGK419	none	355 = W (cagagcagcagcagcagc)
SGK419	35	92	FL	PK	Other	NAK	4q24		AI735391, AW067339, AA705517	none	5932598_100001_1	57	NP_055726	SEQ_ID_25 _5932598_1 00001_1: FL from DKFZp434P 0116_h NM_017593 04-JUL- 2000--	none	3442 = K (cttccagacacacacacacag gacacatg-cdsaa 84374114-14-18-64-K- ag)
SGK125, MYO3A	36	93	FL	PK	Other	NinaC	10p12.32	3145 = Y, 3204 = Y	AW160373, AA478607	none	17000078090910	38	NP_055079	>MYO3A, NM_017433 1Homesapi ensymosinili AmRNA.	none	3145 = Y (atcacacagcagcagcagcagc Y: 3204 = Y (gattcagacagcagcagcagc Y)
SGK445	37	94	partial	PK	Other	PLK	na		none	183 - 200	4000001803382	61	NP_055079 5900002899 3040 (11.5 kb)	>SGK445.	183 taatattaaagacaa 200	

Table 2 - CHR, SNPs, dbEST, Repeats  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Chromosome	SNPs	dbEST_hits	Repeats	Genomic_ID	sort1	Ortholog	Header	Repeats	SNPs
SGK127	38	95	FL	PK	Other	RAF	12q24.21	501=S (s2205786)	none	none	17000062804843	39	>NP_03859 9.1[kinase_s] uppressor_o lncyle l_ras_lmus_7474604CB musculusl-1-	>SGK127 1700006280 4843. lncyl_104192 3. lncyle_1103 660. 1700014002 1687.	none	501=S (ccagggccacS) dbSNP s2205786
SGK008, ANKRD3	39	96	FL	PK	Other	RIP	21q22.3	510=M (s2205786)	none	427-452	11000258161587	5	>SGK009 -	gacctaagcccggaacac tgcct 452	none	510=M
SGK421, STK22A, TSK1	40	97	FL	PK	Other	STK22A	5q31.1	510=M (s2205786); 278=R (s2205786)	none	none	6759077_1_4	59	>SGK421 400001803 622-	>SGK421 400001803 622-	none	(ACAGTGTGCGAATGG CM) dbSNP s2205786; 278=R (gcaagagagctcagf) dbSNP s2205786
SGK047	41	98	partial	PK	Other	STKR	10p11.21	510=M (s2205786)	none	none	11000258262374	18	NP_004320	>SGK047	none	510=M
SGK196	42	99	FL	PK	Other	Unique	na	98=R	none	none	17000062537825	46	T01289	>SGK196 >SGK396 SEQ_ID_31	none	98=R (cgcagagagcccgagaa tgcctR)
SGK396	43	100	FL	PK	Other	Unique	na	AL048857, AL048858, T08068	none	none	5630059_1_4	53	CAB90410 5630059_1 CAA11152_4	>SGK396 SEQ_ID_31 5630059_1 CAA11152_4	none	665=Y (cctgagagtggtgccag gt)
SGK279, PKN	44	101	FL	PK	Other	YMY3 ce	16q22.3	665=Y	none	none	17000097259742	48	BAA36362 ST1887	>PKN-	none	665=Y (cctgagagtggtgccag gt)



Table 2 - CHR, SNPs, dbEST, Repeats  
413406\_1.xls

Gene Name	IDna	IDfaa	FL/Cat	Super-family	Group	Family	Chromosome	SNPs	dbEST_hits	Repeats	Genomic_ID	sort1	Ortholog	Header	Repeats	SNPs
SGK037	45	102	FL	PK	STE	NEK	13q14.12		AA393108	888 - 912	17000030285658	12	P51956, P51954	>SGK037. SGK113. SGK039 13 AUG-2000: FL virtual from gene-wise/ge nscan of celera assembly 1730000362 74838MDK start from HGP 6996171 (positions 144536- 114544) - still missing HPN motif Clem predicted from genescan of celera assembly 1730000362	888 aggggagccacgcccacccg ccgtt 912 485	
SGK060	46	103	FL	PK	STE	NEK	3q22.1	1463=W (441367871); 422=R (ss1855009)	BE388872, AA412114	485 - 488; 1468 - 1487	17000036897142	22	NP_035979	>SGK060 7477585CB 1 : 13-AUG- 2000	1463=W (agggcagcagggg) 1159=R (agggcagcagggg) 422=R (agggcagcagggg) 1467	
SGK080	47	104	FL	PK	STE	NEK	22q11.2		none	none	17000048119006	30	NP_002488	>SGK080 13-AUG- 2000	none	dbSNP ss1855009 (agggcagcagggg) 422=R (agggcagcagggg) dbSNP ss1855009

413406\_1.xls

-130-

Table 3 lists the extent and the boundaries of the kinase catalytic domains. The column headings are: "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Profile\_start", "Profile\_end", "Kinase\_start", "Kinase\_end", and "profile". The contents of the first 7 columns (i.e., "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family") are as described above for Table 1. "Profile Start", "Profile End", "Kinase Start" and "Kinase End" refer to data obtained using a Hidden-Markov Model to define catalytic range boundaries. The profile has a length of 261 amino acids, corresponding to the complete protein kinase catalytic domain. Proteins in which the profile recognizes a full length catalytic domain have a "Profile Start" of 1 and a "Profile End" of 261. Genes which have a partial catalytic domain will have a "Profile Start" of greater than 1 (indicating that the beginning of the kinase domain is missing, and/or a "Profile End" of less than 261 (indicating that the C-terminal end of the kinase domain is missing). The boundaries of the catalytic domain within the overall protein are noted in the "Kinase Start" and "Kinase End" columns. "Profile" indicates whether the complete or "Smith Waterman" (partial). Starting from a multiple sequence alignment of kinase catalytic domains, two hidden Markov models were built. One of them allows for partial matches to the catalytic domain; this is a "local" HMM, similar to Smith-Waterman alignments in sequence matching. The other "complete" model allows matches only to the complete catalytic domain; this is a "global" HMM similar to Needleman-Wunsch alignments in sequence matching. The Smith Waterman local model is more specific, allowing for fragmentary matches to the kinase catalytic domain whereas the global "complete" model is more sensitive, allowing for remote homologue identification.

Table 3 - Kinase Domains  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Profile_start	Profile_end	Kinase_start	Kinase_end	Profile	Comparison	Genomic_ID	Cel Sort
SGK187, CRIK	1	58	FL	1	261	98	361	Complete		4926912_1_1	44
SGK084, GRK7	2	59	FL	1	261	191	454	Complete		11000284009826	23
SGK409, KIAA0303	3	60	FL	1	261	470	743	Complete		6671897_1_1	54
SGK021	4	61	Cat	1	261	23	263	Complete	MGN S	17000028181153	8
SGK410	5	62	FL	1	261	289	557	Complete	KPPVFDENE+V	7018620_2	55
SGK089	6	63	Cat	1	261	158	421	Complete	NFDHF+ILRAIG	11000283338966	26
SGK110	7	64	partial	1	261	68	329	Complete	KG FGKVCIVQK	11000284702500	35
SGK053, CKLJK	8	65	FL	1	261	23	279	Complete	DTKKMYAMKY	17000035790284	20
SGK124	9	66	FL	1	261	73	315	Complete	MNKQKCVVER+E	17000047891899	37
SGK254, CAMKka	10	67	FL	1	261	128	417	Complete	VRNVF+ELQIM	4454511_1_3	47
SGK297, CaMKIb2	11	68	FL	1	261	15	270	Complete	QGLEHPFLVNL	17000113122540	50
SGK411, CaMKII delta2	12	69	FL	1	261	14	272	Complete	WYSFQDEEDM	2828765_1_3	56
SGK027	13	70	FL	1	261	74	325	Complete	FMVVDLLGGD	17000029868654	9
SGK048b	14	71	partial	98	175	1	75	Smith Waterman	ERYHLQQNWHF	11000284253087	16
SGK048c	15	72	partial	106	144	1	39	Smith Waterman	E TVKL+ICEL	11000284253087	17
SGK089	16	73	partial	130	216	1	84	Smith Waterman	+AL+YLQ	11000283986586	32
SGK133, N63965	17	74	FL	1	261	50	301	Complete	IIHRD+KPDNILL	6671894_1_1	40
									DEHGHVHITDF		
									NIA +L +		
									++MAGTKPYMA		
									PE+F G		
									GYS+		
									VDWWSLGVTA		
									YELLRG RPY I		
									S+T EI++ F+		
									V Y S W +		
									MVSLKK+ K		

Table 3 - Kinase Domains  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Profile_start	Profile_end	Kinase_start	Kinase_end	Profile	Comparison	Genomic_ID	Cel Sort
SGK004_MSK	18	75	FL	1	261	27	281	Complete	MVIMSEFSA P G GQGQKPLRV GFYD+ERTLGK GNFAVVKLARH RVTKTQVAIKID KTRLDSSNLEKI YREVQLMKLLN HP+IILKYQVME TKDMLYIVTEFA KNGEMFDYLT NGHLSENEARK KFWQILSAVEY CH+HHVHRDL KTNLLLDGNN DIKLA DFGFGNFYK GEPLSTWCGS PPYAAPEVFEG KEYEGPQDIW SLGWLYLVLC GSLPFDGPNLP TLRQRLVLEGRF RIPFFMSQDCE +LIRRMILVDP +RITIAQIRQHR WM+A+P L	11000257744647	3

Table 3 - Kinase Domains  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Profile_start	Profile_end	Kinase_start	Kinase_end	Profile	Comparison	Genomic_ID	Cel Sort
SGK006	19	76	Cat	1	261	25	261	Complete	+P + +S + F + K+VG YL+G + + +G FAKV EGL + GEK IK I KK+AK+D+YV KN++ E I Q++++PN +QL LET+NSYY++M ELC GNLM++I++K+ LEE + RQ++SA+EH+H G+VH+D + + + L N + G S + STQ GSPAYAAPELL A +KYGPK+DVW +G + MLTGTLPT+E FS+ LYQKMW	11000257829454	4
SGK180, SNRK	20	77	FL	1	261	16	269	Complete	A++ + ++	1700005757785	43
SGK386, MLCKs	21	78	FL	1	261	301	556	Complete		17000140438265	51

Table 3 - Kinase Domains  
413406\_1.xls

Gene Name	IDna	IDaa	FL/Cat	Profile_start	Profile_end	Kinase_start	Kinase_end	Profile	Comparison	Genomic_ID	Cell Sort
SGK003	22	79	FL	1	261	17	299	Complete	M++SGSKAE	11000257741254	2
SGK066	23	80	Cat	1	261	21	281	Complete	+VGGKYKLVRLI	11000283286349	24
SGK041, NKIAMRE	24	81	FL	1	261	4	286	Complete	GGSGFGD+YL I	11000257953829	24
SGK112	25	82	FL	1	261	4	304	Complete	TNGE+VA+KLE	17000035915087	15
SGK038, ERK7	26	83	FL	1	261	13	323	Complete	SGK	17000030278391	36
SGK158	27	84	FL	253	261	522	532	Complete	+HPQLLYESKL	11000283551497	13
SGK429	28	85	FL	1	34	200	315	Smith Waterman	Y	5001549_3	42
SGK152, SUDD	29	86	FL	251	261	347	358	Smith Waterman	ILQGGVGIPH+	11000258249295	60
SGK077	30	87	FL	1	32	484	513	Smith Waterman	WYQEKD	11000283672012	41
SGK093, Wnk3	31	88	FL	1	261	162	420	Complete	NVLVMDLLGPS	11000284063441	29
SGK074	32	89	Cat	1	261	6	342	Complete	LEDLFNFCRR	11000283476699	33
SGK087	33	90	FL	1	261	506	703	Complete	FTMKTVLMLAD	17000036171996	27
SGK295, KIS	34	91	FL	1	200	23	238	Complete	QMSRIEYVHTK	17000113079883	31
SGK419	35	92	FL	1	261	51	341	Complete	NF+HRDIKPDN	5932598_100001_1	49
SGK125, MYO3A	36	93	FL	1	261	21	287	Complete	FLMG	17000078090910_	57
SGK445	37	94	partial	1	71	1	68	Complete	GRHCKLFLIDF	4000001803382	38
SGK127	38	95	FL	1	261	661	922	Complete	GLAKKYRDNRT	17000062804843	61
									RQHIPPYREDK+		39
									LGT		
									RYASINAHLGIE		
									QSRDDMESL		
									GYV		
									MYFNRTSLPW		
									QGL+A		
									TKQKYEKISEK		
									KMSTPVEVLCK		
									GFPAEFAMLYN		

Table 3 - Kinase Domains  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Profile_start	Profile_end	Kinase_start	Kinase_end	Profile	Comparison	Genomic_ID	Cel Sort
									P		
									+DLGTTDVQKK		
									KLVDIVSGDTS		
									KLMKILQPQDV		
									DLALDVGASLL		
									HLAVEAGQEEC		
									AKWILLNNANP		
									NLSNRRRGSTPL		
									HMAVERRVRG		
									WELLARKISV		
									NAKDEDQWTA		
									LHFAAQNGDES		
									STRLLLEKNASV		
									NEVDFEGRTPM		
									HVACQHGQENI		
									VRILLRRGVDVS		
									LQGDRAWLPLH		
									YAAWQGHLPV		
									KLLAKQPGVSV		
									NAQTLDGRTPL		
									HLAAQRGHYRV		
									ARILIDLCSDVN		
									VCSLLAQTPH		
									VAAETGHTSTA		
									RLLHRGAGKE		
									A+TSDGYTALH		
SGK009, ANKRD3	39	96	FL	1	261	22	276	Complete	LAARNGHLATV	11000258161587	6
SGK421, STK22A, TSK1	40	97	FL	1	261	12	272	Complete		6759077 1 4	59
SGK047	41	98	partial	232	261	1	31	Smith Waterman		11000258262374	18
SGK196	42	99	FL	1	88	81	160	Complete		17000062537825	46
SGK396	43	100	FL	1	261	186	425	Complete		5630059 1 4	53
SGK279, PKN	44	101	FL	1	261	53	313	Complete		17000097259742	48
SGK037	45	102	FL	1	261	4	259	Complete		17000030265658	12
SGK060	46	103	FL	1	261	29	287	Complete		17000036897142	22
SGK080	47	104	FL	1	261	8	269	Complete		17000048118006	30



Table 3 - Kinase Domains  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Profile_start	Profile_end	Kinase_start	Kinase_end	Profile	Comparison	Genomic_ID	ICel Sort
SGK002	48	105	FL	1	261	71	368	Complete	MLARRKP+LPA LTINPTIAEGPS PTSEGASEANL VDLQKKLEEL+L DEQQ KRLEAFLTKA KVGELKDDDFE R SEL AGNGGVVTK +HRPSGLIMAR KLIHLEIKPA+R NQIIRE QVLHECNSPYI VGFGAFY D EISICMEHMDG GSLOQ LKEAKRIPEDIL GKVSIAVLRLGLA YLREKHQIMHR +VKPSNILVNSR GEIKLCDFGVS GQLIDSMANSF VGTRSYM+PER LQGTHYSVQS IWSM LSLVELAI RYPPIPPDAKEL EA FG+PVD	11000257702671	1
SGK058	49	106	Cat	1	261	11	274	Complete		17000036877196	21
SGK103	50	107	partial	125	148	1	24	Smith Waterman		11000284272557	34

Table 3 - Kinase Domains  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Profile_start	Profile_end	Kinase_start	Kinase_end	Profile	Comparison	Genomic_ID	Cel Sort
SGK035 SGK075	51	108	partial	1	261	73	324	Complete	PPVIAP PDH KSIYTR+VIDPV APVGDG+VDG AKSLDKQKKK KMTDEEIMEKL RTIVSIG +KKYTRYEKIG QGASGTVFTAT DVALGQ+VAIK QINLQKPKKE LINEILVMKELK NPNIVNFDLSYL VGDELFW+EY LA GSLTDWTETC MDEAQIAVCR E LOALEFLHANQ VIHRDIKSD+VL LGMEGSVKLTD FGFCAQITPEQ SKRST+V TP+WMAPEVVT RKAYGPKV +WSLGIMA EMVEGEPPYLN ENPLRAL	17000030169905 11000283492249	11 28
	52	109	Cat	1	261	15	281	Complete			

Table 3 - Kinase Domains  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Profile_start	Profile_end	Kinase_start	Kinase_end	Profile	Comparison	Genomic_ID	Cal Sort
SGK188, EphA9	53	110	FL	1	261	645	899	Complete	C+ G P +EV LLDSK Q	17000057739161	45
SGK040	54	111	FL	1	261	394	846	Complete	ELGW + PS+GWEEISGV	11000257912897	14
SGK390	55	112	FL	1	13	383	395	Smith Waterman	DEH PIRTYQVCNV++ +Q+NLW+TW+ R	17000112562523	52
SGK007	56	113	Cat	166	261	613	716	Smith Waterman	Q+I+VEL+FTLR DC+SIP GTCKETFN+YY +E++ +R + KIDTIADESFT Q DLG+R +KLNTE+RE+G P+++GF+LAFQ DVGACVALVSV RVY+K+C TV+ LA F P T + +LVEV G+CV +S+ E PPRM+C +GEWLP+G+C +C+AG++ERG IC+AC PGFYK S C+CP HS E+S C C+++Y R+	11000257855312	5
SGK050	57	114	partial	51	89	1	37	Smith Waterman		17000035767558	19

Table 4 describes the results of Smith Waterman similarity searches (Matrix: Pam100; gap open/extension penalties 12/2) of the amino acid sequences against the NCBI database of non-redundant protein sequences (<http://www.ncbi.nlm.nih.gov/Entrez/protein.html>). The

5 column headings are: "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family", "Pscore", "aa\_length", "aa\_ID\_match", "%Identity", "%Similar", "ACC#\_nraa\_match", and "Description". The contents of the first 8 columns (i.e., "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Serial #", "Superfamily", "Group", "Family") are as described above for Table 1. "Pscore" refers to the Smith Waterman probability score. This

10 number approximates the chance that the alignment occurred by chance. Thus, a very low number, such as 2.10E-64, indicates that there is a very significant match between the query and the database target. "aa\_length" refers to the length of the protein in amino acids. "aa\_ID\_match" indicates the number of amino acids that were identical in the alignment. "% Identity" lists the percent of nucleotides that were identical over the aligned region. "%

15 Similarity" lists the percent of amino acids that were similar over the alignment. "ACC#nraa\_match" lists the accession number of the most similar protein in the NCBI database of non-redundant proteins. "Description" contains the name of the most similar protein in the NCBI database of non-redundant proteins.

Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Pscore	aa_length	aa_ID_match	%Identity	%Similar	ACSF_nraa_match	Description	Genomic_ID	Comparison
SGK187, CRK1	1	58	FL	PK	AGC	DMPK	0	2053	1970	98	98	AAC72823.1	Rho/rac-interacting citolon Kinase (Mus musculus)	4926912 1 1	MLKFRYG RNP *A A*EPASRASRLNLFQGGPP MTDOONAS LSREG+LDALF LEECOSPALMKKMS+EV * YSDTUEL+ELQPSA+OFVRSLV GCGPAEVQVVERKATGD+YAM K+MKKQALLAQEQVSTFEERNI LGRSTSPWIPQIQYAFQDKN+LY LWMEYQPGGD LSLNRYEDQDLE+*IQPYLAELI AVHSHV MGYVHRDKPENIL*DRITG IKLYDFGSAKMNSK V+A K+DMCTLDRDANTYTLQRRK+D DS*ELORRRSLALPG OGCAELRQLS +FHSLLCEQPIGRRLRFDFLATV P+*A FLEDVQNWELAECP K S LQ L ATCA P PQ FLSQ +ATKC*AA+T*EER V AKAE M+FLQEQPF+DF+ S FYD+FLOWKLFEMQPVSDKYFT EPRVLGKGGFGEVCAVON+HTG KAYACKLIDKRLKKGGEHMA LLEKELEK+SPFVSLAYAFESK THCLVMSLUNGGLKHHNYNG TRGLMSRVFY+AQ+CG+LHLH LGVTYRD+KPEIN LLDGLGKRLSDQLAVE+* KPTIDRAGTNGYMAPELMA+K SYSYPPDWFAMGCSYEMVAGR TPFKD+KEKYSKEDLK+RT+*DE V F H+NFTEE KDICRLFLAKPQEORL SREK+ODPRKH FF+T+NPRLAAGL+EPFVYDPP SWYAKD+ EIDDFSEVRGVFDDKDKOFF+ F+TGAVP+AWQIEIETGLFEELN DPNRP+G *G+SSKSGVCLL EFDPEEYVYLEAGCHAKEGQ GKTDIPRYISQLGNKPLEMA HLGNYDSGTAEPTDESYSVS NASLKLRRKPRSDPETIKLSNG AYGAVYVRHRESRORFAMKKI KONLRLNQIOQAFVERDILTFAE NPFVSNYCSFEETRHLCHVAE YVEGGDCA TLKNNGLPIQDMA RMYFAETL ALEYLWNGVARD LQPDNLVTSIMGHILTDGLSKY GLMSMTNLTEGHIEKQAEFLD
SGK409, KIA0303	3	60	FL	PK	AGC	MAST	0	2523	2137	100	100	BAA20762.1	KIA0303 (Homo sapiens)	6871897 1 1	
SGK084, GRK7	2	59	FL	PK	AGC	GRK	0	553	466	84	91	AACS5001.1	G protein-coupled receptor kinase GRK7 (Spermophilus tridecemlineatus)	11000294009826	

Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Pscore	aa-length	aa_ID-match	%identity	%Similar	AGC#_naa-match	Description	Genomic_ID	Comparison
SGK021	4	61	Cat	PK	AGC	Mo3C11.1	6.00E-138	327	203	77	84	CAB76566.1	Serine/threonine protein kinase (Mus musculus)	17000028181153	MGN S KPPVFDENE+VNFDFH+ILRAICK GFGKVCIVCK DTKKAYAKYANKKQKVER+EV RNVF+ELQWOGLEHPELVNLWY SFQDEEDMFNVDLLGGDLRY HLOQNVHFETVKL+ICEL ++AL+YIQ IIRRD+KPDNILLDEHGHVHTDFN IA+L++MAGTRPYMAPE+F CGYS+VDWWSLGVATAYELLRG RPYIS+T E1++F+ VYSW+

Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Pscore	aa_length	aa_id_match	%Identity	%Similar	ACCE_nraa_match	Description	Genomic_ID	Comparison
SGK410	5	62	FL	PK	AGC	PKC	0	631	555	95	96	NP_002731.1	Protein kinase C, $\alpha$ isoform [Homo sapiens]	7018620.2	MSH VAGG SGGSHQV VMA YRGDMITHEPS-SFEG N VROMCSFONEQLFTMAKWIIEEG DPCTVSSQLELEAF LYELAKOSELHVFQVPERPG NPKCF EDKSYRRGAR WRKLYCANGHTFQAK FNRRA CAICTDRWGLG OGYKCHCKLLVKKCKHL-TIEC GRKSLP EP-MPNDOSMSHDAQTVIPY NPSSHESLDVGEKEAMATRE SGKASSSLGLODFLL-VIGRGS EDM L L * L * * V * Y E VR LG-G YG-VLV * * GT * ALK * K * T L * F L E * S L * I * * E * Y F E GDL IP-VGLP- VRC QL AL * * * H * R LV-RD-VREN-L D CRR KL DFG TR G * IPATAP-C FG * * D WA GYL-FCKTG FV * * * AD P * E * * WO P P W * DTH * KLG G * * VLL * G * VALK * R * D * * F * E * * S * PG * * * T * * T Y F D E A P G L * * * E * G * P E * * VKR * Q * ALDF-H RLVHD * * H * D * N * L * D * C * L D G T * G S P * * P P E L C L P L L * * D W G * L * F * A T F P W * A * N * * * F W * * P P W * * F A * A * LL P * R
SGK069	6	63	Cat	PK	AGC	Unique	2.10E-64	572	116	42	58	BAA36362.1	Serine/threonine protein kinase [Rattus norvegicus]	11000283338966	MARENGESSSWKQAEIDKGF EFKELTGTGAFSEVLAEEKATG KLFVAKCIPKAKLKGESSIEIEA VLRKIKHNVALEIDIESPNHLY VMDLVSGGELFDRIVEKGFTEK QASTLURQVLDVYVYLRIGGVH IDLKPEALLYVSDDESKWISDF GLSKUEGEGGVASTACTGPGVY APEVLAQKPYKSAVDCHSIGVIA YELLOGTPPTFDENSKLFEQIK IAETFDSPYWDSDSADKF NRATPDAPAGSLSRRGRIEDD NLDTERPVOKRRSGGPPRPP CLPLSPPTAPDRATATAASRL GPYLLLEPEEGGRAYQALHCT GTEYTCVKVYPOEALAVLEPYAR LPPHKKHVARPTVLAGTQLLYAF FTRTHGDHSLVRSRRIPEPEA AVLFRQMATALAHCHQHGVLVR DLKLCRFVADREKRLVLENLE DSCVATGPDSDLWQKACPAVY GPELSSRASYSCKAOWSLGV ALFTML
SGK110	7	64	partial	PK	AGC	Unique	1.10E-65	373	112	40	58	S71887	pk9.7 gasinula-specific PK [Xenopus laevis]	11000284702500	
SGK053, CLKIK	8	65	FL	PK	CAMK	AMPK	1.60E-245	357	357	100	100	NP_065130.1	CamK-like protein kinase [Homo sapiens]	17000035790284	
SGK124	9	66	FL	PK	CAMK	AMPK	5.10E-252	358	358	100	100	CAB81634.1	Novel protein kinase [Homo sapiens]	17000047891899	

Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	ID#na	ID#na	FL/Cat	Super-family	Group	Family	Score	aa-length	aa_ID-match	%Identity	%Similar	ACC#_naaa-match	Description	Genomic_ID	Comparison
SGK254, CaMKK3	10	67	FL	PK	CAMK	CAMK	2.0E-323	513	470	91	94	A57156	Ca2+/calmodulin-dep. PK IV [Rattus norvegicus]	4454511.1_3	ME PAVCGDPRALVELVERVAM V HLEEA GREP NGVDP PRRAASVPOS SR P RPSLSARK SLGERPA S LEAQ GPY*IGPASH*SPRAWRRPTIES HHVAISD EDCVQLAQYKLOSEIGK GAYGVRLAYNE EDRHYAMKVLKXKKLLKQYGF RRPPRGSA QGGPAKQLPLRLERYQELAKKL DHVNVKLVLDLDDPAEDNLVY DLRKGPMVEVPCDKP
SGK297, CaMKIb2	11	68	FL	PK	CAMK	CAMK	1.40E-236	343	341	99	100	AAF74509.1	Ca2+/Calmodulin-dependent protein kinase I [Homo sapiens]	17000113122540	MLLKHTEDISVYERERLGSQ AFSEVLACDERGSAHLVAKCIP KKALRGKEALYENEAVLRISHP NVALEDVHESPSPHLYAMELV GGELFRIMERGSGYTEKASHLV GVALCAVSLHSLGVAHDLKPE NLVYATPESDKIHWVSDFGLSKQ AGNMLGTACGTGTYAPELLEQ KPTGKAYDVWALGVISVILCOY PPFYDESDELPSQURASYEFD SPFW DISESAKDFIRHLERDP
SGK411, CaMKII del2	12	69	FL	PK	CAMK	CAMK	0	499	499	100	100	AAD20442.1	Multifunctional CAMK II del22 [Homo sapiens]	2828765.1_3	MASTTCTRTFTDEYQVFEELKGG AFSWRRCKMKTPTGGVAAKINT KKLSARDHOKLEAREICRLUGH PNVRLHDSISEEGFHYLVFDLVT GGELFEDVAREYSEADASHCI QQILES*VNHCHLNGVHDLKPE NLLASKSKGA/VKLDAGLAIEV QGQQAWFGAGTPTGYSLEVL KMDPYGKPYDAMWACGVLYLLV GYPPFWDQDQHLRYQOIKAGAY DFPSPWDVITVTEAKLNKMLTI ** LG*RGFYR* E*G*GNFS*VKLG*H LTKEVA*KK*DK K*+D K O*LLSREI *ME***HPNI*+L*EVETL***HL V*EYA GGEI*+ GKL*E **K *F*OVSAYNMH *H*RD*KAENV**S VK* DFGS*+*ML VCGSPPTAAPELF*+D Y G VD*WALGVLL*FM* G PFA*ETV *K I G Y*P*VS L*+L* *R ID
SGK027	13	70	FL	PK	CAMK	EMK	7.70E-101	436	147	56	76	T22427	F49C5.4 - [Caenorhabditis elegans]	17000029868654	**E*EA* FCH*AA*YC K**H D*+*LL E NIKI DFGFS F * FCG PYAELF
SGK046b	14	71	partial	PK	CAMK	EMK	1.40E-14	75	35	47	61	AAC33487.1	R31237_1, partial CDS [Homo sapiens]	11000284253087	
SGK046c	15	72	partial	PK	CAMK	EMK	8.40E-07	39	21	54	67	AAC33487.1	R31237_1, partial CDS [Homo sapiens]	11000284253087	
SGK089	16	73	partial	PK	CAMK	EMK	2.40E-19	84	39	46	71	AA97437.1	Serine/threonine kinase [Caenorhabditis elegans]	11000283986586	FRQI*S*+YC KNI H*+L*+ E NIKI DF N* D** NIKIA**GFS*TF*+* TF G* PYA PELF G Y P **WWSLG*H*+V*G*JPF ON K**++



Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	IDna	IDaa	FL/Cat	Super-family	Group	Family	Pscore	aa_length	aa_ID_match	%Identity	%Similar	ACC# naa_match	Description	Genomic_ID	Comparison
SGK133, N83965	17	74	FL	PK	CAMK	EMK	8.2e-318	794	481	75	83	CAA07196.1	Putative serine/threonine protein kinase [Homo sapiens]	6671894_1_1	LIENPHVYDTHVYVKKRYLYVC ENVSGLGFLDYLVKKGLTPKEA RKFRQI+SALDFCHS+SIQRDL KPEALLDEKNIRIADFGMASLO VGDSLLETSCGSPHYACPEV+G EKYDGR+AD+WCGVILFALLVG ALPFDDNULRLEKVKROVFH PHFPPDCOSILRGM EV+ +RL+LE IQKH WY+GCK+EP+P +P R+V +RSLPs ++DPDV+SM SLGCFROR +L +L S WVMSFSDNPAGCGCGQKPL RVGFDIERTLCKGNFAVVKAR HRVYKTQVAKIDKTRLDSSNLE KITREYQLAKLUNPHIKLYQVM ETKMLYVTEPAKNGEMFDYLT SNHLSSENEARKKFWQLSAVEY CHDHVHRDLKATKLLDGNMD IKLAGTDFGFGNPKSGEPLST WCGSPYAAPEVEFEGKEGPGQ LDWISGLVLYLVCGSLPFGP NLPTLRQRYLGRFRIFPFMSQD CE
SGK004, MSK	18	75	FL	PK	CAMK	EMK-1	0	788	784	100	100	P57059	SNF1LK [Homo sapiens]	11000257744647	+P+ +S+ F+ K+VG YL+G++ +G FAKV EQL+ GEK IKI KK+AK+D+VY KN++E I Q++++PN+QL LET+NSYY++MELC GHLN++it++K+L E+ RC++SA+EH+H G+VH+D + +L N+ G+S+ STO GSPAYAPPELLA+KYGK+QVW AG +MLGTGTPFTLE FS+ LYONKXV A++++
SGK006	19	76	Cat	PK	CAMK	EMK-1	1.50E-62	262	121	45	66	NP_056570.1	Hormonally upregulated Neu- associated kinase [Mus musculus]	11000257829454	MAGFKRGYQKAGLYDLDKTLG RGHPA +RRVFTGERAVKVKIDTKLDTLA TGHLPQEVRCMKLVQHPNVRLY EVIDTQTKLYLLELGO GDMFDYMKHHEEGNEOL KKYFAQVHAISYCHKJWVHRDL KPENVFFEKQGLVKLTDGFSN KFGPKGLTTSCGSLAYSAPELL GOEYDAPAVQWISLVLFMLVC GPPFQENQSELTIMIDCKY TVPSHV/SKCKULTRMLQRDPK
SGK180, SNRK	20	77	FL	PK	CAMK	EMK-1	0	765	757	99	99	AAF86944.1	HSNFRK [Homo sapiens]	17000057577785	MA TENGAVELGIONPST DKAPKPTGERPLAAGKDPGP DPKAPDPPTLAKDAKASEKG DGLAOPSTISSQKPKGEGDRG GGPAEGSAGPPALPOOTATPE TSVKKPAEGOGAGSGSQPKPR VCKKAMEGQAAARRGSPLEUS PQCPAIISSSEKLANQPTSEASE LTFEGVPMHSPTDPAFAACE GNILASQKEVGEKTPGOAGQ AKMGDTSRGIEFQANVPSEKSE
SGK386, MLCKs	21	78	FL	PK	CAMK	MLCK	0	612	596	97	97	CAC10006.1	MYLK (myosin, light polypeptide kinase) [Homo sapiens]	17000140438265	

Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Pscore	aa_length	aa_ID_match	%Identity	%Similar	ACC#_ntaa-match	Description	Genomic_ID	Comparison
															N+SGSKAE +VGGKYLVRKIGSGSFGD+YL I TNGE+VA+KLESQK +HPQLYESKLY ILOGGVGIPH+ WYGOEKD NVLVNDLLGPSLEDFNFCSSRF TMKTYLALADQMISREYVHTKN F+HRDQKPNFLMG GRHCKNLFIDFGLAKKYRDNRT ROHIPYRECK+L GT RYASIAHLGIEQSRDDMESLG VYMYNRTSLPWQGLA+ TKKCKYKESKKNSTPVEVLCK GEPAE 11000257741254
SGK003	22	79	FL	PK	CKI	CKI	1.40E-215	337	304	91	95	NP_001883.2	Casein kinase 1, alpha 1 [Homo sapiens]		KIGGGGGEIY+A D+ E VAKYGS+ KQVLMEEVAVL+LOG +Y GR+ **MS G+NLADRR + F +ST+R+G OIL I IHS+GLHRD+KPSNFAMGR T R YMLDFGLARQ+N+ G+R PR+ AGFRGTVRYA+ AH+N+EMGR DDLWSLYML EF+ GGLPWRKIKK+VG +KE D +L P E +F+ H+
SGK066	23	80	Cal	PK	CKI	CKI	5.40E-106	400	166	53	68	T24262	R90.1 [Caenorhabditis elegans]	11000283296349	MEWYETLGKVGSGSYGTVMCK HKNTGQIVAKIFYERPEQSVNKI AMREIKFLQKHENLVNLEVFR QKKIHLVFEFDHTYDLEOHYC HGLESKRLRYLFOILRAIDYLS NNIHRDIKPENILVSGSITLDC FGFARTLAAPGDYDYATRWY RAPELVLDKTSYKPKVDIWLGC MIEMATGNPYLPSSSDLLHKK VLKVGNSPLHONIFSKSPFAGV VLPQVQHPKRNKYPKL
SGK041, NKIAMRE	24	81	FL	PK	CMGC	CDK	1.0E-319	581	454	99	100	NP_057592.1	NKIAMRE [Homo sapiens]	11000257953829	MEKYEK+ K GEGSYGVFKCRN+ +GQ+VA+KCF+ESEDPPV+KKIAL REIRMLKQLKHPNLVNL+EVFR KR+HLVFEYCOHT+L+EL+R GV+ **KS+ WOTLQ+MECH NHCNRD+KGENILTK +IK+CDGFEA+L+ P D YTDVATRWY+PELLVGDGTQY G VD+WAIGCVFAELL+G PLWFGKSDVDLLYI+TLG LIPRQ +F+N +F G+ M+KY +*IG+GFGAT L K D +H VIKEN ENG L+VMA+YC+GGDL KRIN Q+G LF EDOIL WFOIL LKH+HDKLHRDIK+QNIFL+K+G V
SGK112	25	82	FL	PK	CMGC	CDK	8.70E-151	380	224	61	74	NP_004187.1	CDC2-related kinase [Homo sapiens]	17000035915087	+LGDGFIARVJN+*ELARTCIGT PYLSPFIC+KPYNNK+DIW+LG CVLYELCTLQK FE N+LVLI F P+SP +S +L SL+SQLF+ +PRDRPS+NSIL+
SGK038, ERK7	26	83	FL	PK	CMGC	MAPK	1.80E-105	371	167	59	72	P51954	NEK1 (NIMA-RELATED PROTEIN KINASE 1) [Mus musculus]	17000030278351	

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Pscore	aa_length	aa_ID_match	%Identity	%Similar	ACCH_nraa_match	Description	Genomic_ID	Comparison
															MSIQDQAFRIPLUGKPEKVRQS ADP*PLQIMAKTKUNDLGPWY RDKLEYFEERPFMAASIGQVHLA RMKGGREVAHGQYPCVQAQSN SDVNNLMAVLNNSNLLPEGLFP EHLIDVLRRELAECQYQREAA ARKEFDLLKGHPFFVYPEIVDEL CSPPHATTELVSQFPDQAEGLS QEIRNEICYNVLCLCLRELFEHP MOTDPNNSNFYDPQDHKVALL DFGATREYDRSFDTLYQIIRAAA DRGR
SSGK158	27	84	FL	PK	Microbial PK	ABC1	3.30E-257	751	368	100	100	NP_054632.1	Hypothetical protein [Homo sapiens]	11000283551497	IFVNFVIAQDQGRNLLVGQANG LESSQEAQLQADICDTLVAVP SSLCPLRLVLDAGNAELQAPDL RNFRALFMAVWGQGGRIAEILI LHARASECRDVEGFKETEMAL VTOARAKTITLKKVSSLSVYF KLIMTKYKALESHFASFPAAWY EGLGRSLDPRLDLKAAARPLLTG VCTV
SSGK429	28	85	FL	PK	Microbial PK	ABC1	6.00E-122	626	191	100	100	AAD43192.1	Putative human protein [Homo sapiens]	5001549_3	MDLYGVTSPEFGTAAAWGFSQC PWAIPQNTISGLADNMSQELAK ELQLEEEAAVPEVAVAEGPFT GENIDTSSQLAQILQMEYDR EYDQALRREEKFGNGSKVSI ENRYKXVHYEDSDSDSEVDVQ DTRDDPYRPAKVPPTPKKGFIGK GKDIITTHDEVCCGRKANTARME NFAPFQVGDGIGNDLKLSNHVF NALVQAHAYSEERSARLHEKKE HSTAEKAVOPKTRLLAYKAVNS GALETTIG
SSGK152_SUDD	29	86	FL	PK	Microbial PK	R101	0	519	519	100	100	NP_003822.1	sudO (suppressor of bimD6, Aspergillus nidulans) homolog [Homo sapiens]	11000258249295	MA*P*G*RLRYA* R*OR*G A*QWF*P*+R*F*SS S*+* DDPDQDIPFGS V*RRR P*G +++*L*TP*+* LR RPPQKCSSTPC LG PPEPA* GL D*SV*Q*RD*ELG SASLFFS AS PG*P*P LA SAVP HLP SL S*P*Q*ATG R*+* AS RS L*F*H*NSG P*+*P*G*+*P*+*+*+*+*
SSGK077	30	87	FL	PK	Other	C26C2 ce	8.50E-306	798	492	62	73	BAB00640.1	Hapsin [Mus musculus]	11000283672012	++EP*GG* E ER**Q** E**ET*AV S DGR*LFQDIEGRGSFTY*GLD T*TTVEVAWCELG RKL***ERDRP EE EMLXGLQHPNVRFYDSW+S ++G+
SSGK093_Wm3	31	88	FL	PK	Other	C26C2 ce	2.10E-203	1513	321	60	73	AAF31483.1	Kinase deficient protein KDP [Homo sapiens]	11000284063441	CIVLVLTSLTSGTKLYL*RF* MK*+VL* W*QRL*GL FLH*+R PPI*HRDLKDNH*ITGFTGSVY GDLGLATKRAFSKSYGVTP

## Table 4

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Pscore	aa_length	aa_ID-match	%identity	%Similar	ACCE_nraa-match	Description	Genomic_ID	Comparison
SGK074	32	89	Cal	PK	Other	DYRK	2.00E-107	355	176	50	67	AAD41593.1	Myak-S [Mus musculus]	11000283476699	S T Y + + E L G + C T G + V A K W + R S T E + V A I K L N R + E + + L + + + E + + R E F L V F E + L E O N L + + F K + N F + P L P + + I R + Q V T A L + L K L + I H A D L K B E N I L V D R P + R I K K I V D F S S A S S V Y + G S E + Y R O P E N L G L P F C E + D + W S L G C V + A E L L G W P L P G + T O O + R T + T O G L P + L A K E F R + P + +
SGK087	33	90	FL	PK	Other	DYRK	6.20E-183	703	267	100	100	AA591393.1	DYRK4 [Homo sapiens]	17000036171996	E N O A N K V P A S E L K A S E I F P H P S I K T O D P K A E E R S P K K Q V T L T M A E A L A F P K Q L S P T E G S E I L G T A E L W F L G L E A K A L D T A P E K F S K T S F D D E H G P T L A Y L V H D H A Y R E V L E T I G K G S P G Q V A K C L D H K N E L V A L K I R R K R F P Q O A L M E L K I E A L R K Q D K N T Y N V H K K O F F Y F R N H F C T P E L L G N L Y E L K N N N F Q G F S L S W R P T L S V L K C L M L S V E K I H C D L P E N I L V K O G Q A S V K N A G S G C A W G E P P R F L E A F G R L W Q V Q S R L G S G S S A S V Y R V R C C G
SGK285_KIS	34	91	FL	PK	Other	DYRK	1.10E-292	419	415	99	100	NP_058989.1	Kinase interacting with statmin [Rattus norvegicus]	17000113079883	P G S P P G A L Q P L P Q T T G A A A S A A E Y G F R K E A A L E Q L Q G H R N V T L Y G V F T H F S P N Y S R C L L E L L D V S Y S E L L + Y S S H G C S A M W M I Q H C A R D V L E A L F H H E G Y Y H A D L K P R N I L W S A E N E C F K L I D F L S F K E G N O D V Y I Q T D G Y R A P E A E L Q N C L A Q A G L O S D T E C T S A V O L W S L O I L E M F S G K M L K H T Y R S Q E W K A N S S
SGK419	35	92	FL	PK	Other	NAK	5.00E-252	661	355	100	100	CAB70893.1	Hypothetical protein [Homo sapiens]	5932598_1000011	M K K F S R M P K S E G S G G G A A G G A G G A G A G C G S G S S Y G V R V F A V G R H Q V T L E E S A E G G F S T V F L V R T H G G R C A L K R M Y N N N P D L N Y C K R E T T M K E L S G H N I V G L D C A V N S I S D N Y W E V L L M E Y Q R A G O V W N A N K K L Q T G F T E P V L Q I F C D T C E A V A R L H Q C K T P I I R D L K V E N I L N D G N V L C D F G S A T N K F L N P Q I Q G V N W E E E I K Y T T L S Y R A P E M I N L Y G G K P I T T A D I W A L G C L M P F I G T I F D N F D S P D Y V E I T
SGK125_MYO3A	36	93	FL	PK	Other	NinaC	0	1615	1615	100	100	NP_059129.1	Myosin IIIA [Homo sapiens]	17000078060910	T I G K Y T G K V K V L N K N G O K A A V K I L D P H D I D E E I E A E Y N I L K A L S D H P N A V R F Y G I V K D K V G D K L W L V I L C S G S S T D L V K G F Q R G E R M S E P A L I L H A L M G L Q H L H N K T P R D Y K G N I L L T T E G G V K L V D F G V S A C L Y T S T R H R R N T S V G T P F W M A P E V I A C E Q Q L D T Y D A R C D T V S L G T A I E L G D C D P L A D L H P M A L F K P R N P P K L R O P E

Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Pscore	aa-length	aa_ID-match	%identity	%Similar	ACC#_naa-match	Description	Genomic_ID	Comparison
SGK445	37	94	partial	PK	Other	PLK	9.90E-25	68	47	65	81	AAC37649.1	STK [Mus musculus]	4000001803382	FKVGNL GKGSGVY* ESHTGLE***MI KKA* *QRV NE-K-HCOLKHP LELY YE-D*
SGK127	38	95	FL	PK	Other	RAF	8.60E-254	945	446	47	61	NP_038599.1	Kinase suppressor of ras [Mus musculus]	170000282804843	**ALQOC *Q-ID-SI*L GLRTKC* SNDLTO+EIRLE+KLWY +Q K V +R AEL*+P* W I+VR EN+*EIP +L+L LEM +* E + +*GA+EEC+RL +LCLRV GG + W+T+ +*+PP S* S P TP +*+VAP PP S
SGK009, ANKRD3	39	96	FL	PK	Other	RIP	0	784	784	100	100	NP_065690.1	Ankyrin repeat domain 3 [Homo sapiens]	11000258161587	REGDGTFRVLLALRYPLAGEP TGKRVGSGPGDYTRYRHH WKTWAKCSPSLAYDORERME LLEAKGMEMAFRYLPVYGR EPVGLVMEYMETGSLKLASEP LPWDRFRHETAVGNMFQCM APPLLDLQKPNILLDAHYKIS DFGLAKCNGLSHSLNDGLF GTIAYPPERIREKSLFOTRHDV YSFAVINGVLTOKGPFADENIL HIMVYVKGHRPELPVPCRRPR AC
SGK421, STK22A, TSK1	40	97	FL	PK	Other	STK22A	6.40E-209	367	307	84	90	NP_033461.1	STK 22A (spemogenesis associated) [Mus musculus]	6759077_1_4	NDGAAVLRGRG+*GIRLGEQSY AKVKSAYSERLKNVA+KIDRKK AP+DFLEKFLPREILAMLNH SI+KTYEIF TS GKYYMMEL VQGDLE IKTRGAL ED+ARKKFHLS AIKYCHOLDVVRDLK +NLLDKDFNKLDF FSKRQROSGR+LSKTFGCS AYAAPEVLOGIPYQPYQWKL GVLYHAGCSNPYQDSNKK LRQKEHVHVFPRSKALTGEQKD LY ALQD
SGK047	41	98	partial	PK	Other	STKR	0.265466	31	11	46	67	NP_004320.1	Bone morphogenetic protein receptor, type IA [Homo sapiens]	11000258262374	R +*+S CW PA+RL AL+I MEKQPNRSRRGDPREVPVAVG LUUMALMNTLLYCLDHFIAPR STVDPTCHPTGHRIGQWKNCS PWLSCZELRTEVRQLKRVGGA VKRYFLSEWKEHVALSQTSL MKDDFLHGLQMLKSLGTWVVT LLGYCEDONTMLTEYHRLGSLN LEETLALSQYQNVNWHRLRLA MDYYSINYUHKSPGVTRVNCDS NDLPKTLSDYLLTSNLSLANDL ALPLVNHSSGNLWCGHRELHG DFVA
SGK196	42	99	FL	PK	Other	Unique	2.70E-248	350	350	100	100	BAB15623.1	Unnamed protein product [Homo sapiens]	17000062537825	NA+*+PLN E L+M GVA+GL +H +*HG+*+NVF + G+* + + L +PE+ +*+P SD+Y+*GL+L + K+GL V
SGK396	43	100	FL	PK	Other	Unique	3.00E-09	472	34	28	50	BAB11570.1	Receptor-like protein kinase [Arabidopsis thaliana]	5630059_1_4	

Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Pscore	aa-length	aa-ID-match	%identity	%Similar	ACC#_nraa-match	Description	Genomic_ID	Comparison
SGK279_PKN	44	101	FL	PK	Other	YVY3_ce	4.10E-276	424	400	94	95	BAA35352.1	Serine/threonine protein kinase [Rattus norvegicus]	17000097259742	MSVGCPEEP SLCCGPG AP PGAGVPLLTEDMOALTTLTAAS DVKTKHVELVRELGGTGYGADLY YKGTGTMALVFNKSKTKUNF LREVTINSLSSSPFIKIVYVFE TE*CTVFAQETAPAGDLDIIPPO VGLPDTVKRCVQGLGLDFM H KQLVHRDKPEMLVDFRECRV KLADFGMTRRVRGVRVSGTI YTAPEVCQAGRADG AVDTGVDVWVAFGLVFCATGNF PWEEA M*RY **IG*G*FGKALK'D *H VKEIN *M *E++ S++EV *L MIDPNIV * SP*ENG L*VM*YC*GGDL KRIN Q*G LF EDQIL WFOQIL LKH*HDKLHRDK*QNIFL*K*G V *LGDGFIARV/N*++ELARTIGT PVLSPKIC*AKPNYK*DIWALG C*LYELCTLKHFE N++ LVLI F*P*SP *S*L SL*SQLE* *PRDPS*NSL*F* I K*LR*E KGSGLATLTIGTPHYMSPEAL KGGDTKSDNSLACL YEMCOAM NHAPAGSNPLSVKIVEGDTPL PERYPKELUNAMESMLNKP*SLR PSAIELKIPYLDQDLNACRY EMTLEDKLDCKOKEAAHINAMQ KRIHLOTLRALSEVQIOMTPREEM RLRKLQAADEKARKLVKEEY EENSRMOELSRNFQQLSVQV LHEKTHLKGMEKEEOPGRLS CSQDEDEERWQGREESDEP TLEN MPSRAE*YEVLYTGTGS GRCKI*RKSDGKLVKXEL YGSMTAEKQMLVSEVLL ELK*PNIV YYDRIDRTNTLYVMEYCE GDLASVITKTKEROVLDDEEFL RV TQITLALK CHRRSDG HTV* RDLKPA*YELQCKONKQGLD GLARLHDTSEAKTNGTPTTMM SPEQ NMSYNEK DIWLSGGL YE MUPRTAFSQELACKRGKGF RRYRYSDENELIRMLX
SGK037	45	102	FL	PK	STE	NEK	4.10E-56	649	176	65	83	P51954	NEK1 [Mus musculus]	17000030265658	
SGK080	46	103	FL	PK	STE	NEK	1.3e-322	645	463	100	100	BAB15872.1	Unnamed protein product [Homo sapiens]	170000306897142	
SGK080	47	104	FL	PK	STE	NEK	3.80E-296	446	404	90	93	NP_002488.1	NIMA (never in mitosis gene a)-related kinase [Homo sapiens]	17000048119006	

Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	ID#na	ID#aa	Super-family	Group	Family	Pscore	aa_length	aa_ID_match	%identity	%Similar	ACCh_nraa_match	Description	Genomic_ID	Comparison
SGK002	48	105	FL	STE	STE11	3.10E-248	399	369	92	94	P36507	MEK2 [Homo sapiens]	11000257702671	MLSRPPTDPTATNTTAEQSP TSCGASEANVDLOKALEELD EQQ KRLEAFLTKAKYGEKDDOFER SEL AGAGGVYTVQHRPSPGLIMARK LHLEKPAIRNQIRE OVLHECASPYPVGYGAFY D EISQEMHMDGSLDQ LKEAKRIPE*ILGKVSIAVLRLAY LREKHQIMHR*VPSNLSVSRG EIKLCDFGVSGQLDSMANSFVG TRSYNAPEERLOGTHYSVGS INSM L MEEVPGSSISINREFRDELNVF CYTHQLOGVYVLEHCARD IKGNVVALMPTGIKILDFGQARR LAWAGLNGTHSDALKSMRGTPY WM PEYNESGYGRKSDWSGCTVF EMATGKPPASMDRMAAMFYIG AHRGLMPLLPDHFSEMAADPVR MCLTR
SGK058	49	105	Cal	STE	STE11	3.00E-116	278	167	99	99	BAB15538.1	Unnamed protein product [Homo sapiens]	17000036877198	DL PON+LIDK++ L FLGAR L
SGK103	50	107	partial	STE	STE11	0.008942	28	16	57	68	CAA39285.1	Fused [Drosophila melanogaster]	11000284272557	PPVIAP PDH KSIYTR+VIDPV APVGD+VDG AKSLDKKK KMTDEIMEKLRIVSIC +KKTRYENGQASGVFTIATD VALGQ+YAKQINLQPKRELIIN ELVAKELAPNINFLDSTLVGD ELFVY+EYLA GSLTDWYETCMDEQAIAVCRE LQALEFLHANOVRHRIKSD+VLL GMEGSYKLTDFGCAQITPEQS RST+V TP+MMAPVWTRKAYGPKV +WSLGLMA
SGK035	51	108	partial	STE	STE20	7.60E-212	348	318	92	95	Q13177	PAK-2 [Homo sapiens]	17000030169905	++ PDP+DTWEI ETICKGTGYKV+KV AK++G AAVKILDP+D+DEEIEAYML+L +HPNVV+FYG++K D G +LWLVLKLC+GGSVT+LVKG L G+R+ E +H YIL AL+GLOHJHNN+ IHRDVKGNNILLTTEGGVKLVDF VSAQLTSTR RRNTSVGTFVWMAPEVACEQQ D++YDARCD WSLGTALGQGPRL D+HP++ LFGKRNPPPL PE W EFN FIS+
SGK075	52	109	Cal	STE	STE20	4.00E-158	1106	227	67	80	NP 059129.1	Myosin IIIA [Homo sapiens]	11000283492249	

Table 4  
Smith Waterman  
413406\_1.xls

Gene Name	ID#na	ID#aa	FL/Cat	Super-family	Group	Family	Pscore	aa_length	aa_ID_match	%Identity	%Similar	ACCD_nraa_match	Description	Genomic_ID	Comparison
SGK188, EphA9	53	110	FL	PK	TK	RTK-11	0	1009	538	54	71	Q61772	Ephrin Type-A receptor 7 (Mus musculus)	17000057739161	LC LU G AAEVHLDSKA Q ELW+P+GWEEISG+DE+ PIRTYQVC V+EPNQ+HNL+T WIS+G GRIFVEL+FLRDC+S+PG GTCKETFN+YY ET+ D GR ++ KIDTIAADESFQGDGGERKMKL NTEVREIGPLS++GF+LAFQDVG AC+ALVSV+VYK+C V LA FP T SFS+LVEVGTCA+ +E E + PRMHC A+GEVLVP+G+C GA G+GFG+V+Y G+VAVK ++ E V+ LH+L+LL ++ +VME SKG+L L+ +L TQ + ALHVA+G+YLS +++RDL N+L+ ++AK+D+G+A+ RG++S +APE +N ++ +DVA+SEG+L++G ++ +E ++P P+V L+ C+ P ++L NAGAG DHRPP GGAAG A +A AGRGDSSEAEQ GPKLRKYSTSGQRTKTSKEG QLKQTSSFORWRYKRLGR TLYAKOSLUFEDVLSASVA EASTKNNKSPITITPRLMLCA ENRGMEDWISLSKYQ+REPY EVAQNVHFPSOMHWYAGSH ARPTCAVCRSLSGVTSKGLS CEVCKPAHRCVATRNCKW ITLASIGKDIEDGAVMPHQW LEGNL NA+ E+P+ G E H A H +LVL LF +L+TCL+AKLTGPF PWNIHPFS+ORLGAGLQIA+DK +NSE V GN+ TYTN+ C+ KESLA F +QVQ E IS L GPACPEAAE +IG VL+ESLQYLQW++IG+FGG S SSW +VELW+ VE+EL+ HF ITA +RY++ LLOE L +SS+ARVIL CSSEDA
SGK040	54	111	FL	PK	TK	Unique	4.50E-26	909	76	30	52	P41243	BATK (Rattus norvegicus)	11000257912897	++L BR ++L NAGAG DHRPP GGAAG A +A AGRGDSSEAEQ GPKLRKYSTSGQRTKTSKEG QLKQTSSFORWRYKRLGR TLYAKOSLUFEDVLSASVA EASTKNNKSPITITPRLMLCA ENRGMEDWISLSKYQ+REPY EVAQNVHFPSOMHWYAGSH ARPTCAVCRSLSGVTSKGLS CEVCKPAHRCVATRNCKW ITLASIGKDIEDGAVMPHQW LEGNL NA+ E+P+ G E H A H +LVL LF +L+TCL+AKLTGPF PWNIHPFS+ORLGAGLQIA+DK +NSE V GN+ TYTN+ C+ KESLA F +QVQ E IS L GPACPEAAE +IG VL+ESLQYLQW++IG+FGG S SSW +VELW+ VE+EL+ HF ITA +RY++ LLOE L +SS+ARVIL CSSEDA
SGK390	55	112	FL	PK-like	DAG kin	DAG kin	0	1164	1063	91	94	Q64398	Diacylglycerol kinase eta (Mesocricetus auratus)	17000112582523	++L BR ++L NAGAG DHRPP GGAAG A +A AGRGDSSEAEQ GPKLRKYSTSGQRTKTSKEG QLKQTSSFORWRYKRLGR TLYAKOSLUFEDVLSASVA EASTKNNKSPITITPRLMLCA ENRGMEDWISLSKYQ+REPY EVAQNVHFPSOMHWYAGSH ARPTCAVCRSLSGVTSKGLS CEVCKPAHRCVATRNCKW ITLASIGKDIEDGAVMPHQW LEGNL NA+ E+P+ G E H A H +LVL LF +L+TCL+AKLTGPF PWNIHPFS+ORLGAGLQIA+DK +NSE V GN+ TYTN+ C+ KESLA F +QVQ E IS L GPACPEAAE +IG VL+ESLQYLQW++IG+FGG S SSW +VELW+ VE+EL+ HF ITA +RY++ LLOE L +SS+ARVIL CSSEDA
SGK007	56	113	Cat	PK-like	GCyc	GCyc	5.20E-265	884	499	51	62	T42260	Guanylate cyclase (Rattus norvegicus)	11000257855312	++L BR ++L NAGAG DHRPP GGAAG A +A AGRGDSSEAEQ GPKLRKYSTSGQRTKTSKEG QLKQTSSFORWRYKRLGR TLYAKOSLUFEDVLSASVA EASTKNNKSPITITPRLMLCA ENRGMEDWISLSKYQ+REPY EVAQNVHFPSOMHWYAGSH ARPTCAVCRSLSGVTSKGLS CEVCKPAHRCVATRNCKW ITLASIGKDIEDGAVMPHQW LEGNL NA+ E+P+ G E H A H +LVL LF +L+TCL+AKLTGPF PWNIHPFS+ORLGAGLQIA+DK +NSE V GN+ TYTN+ C+ KESLA F +QVQ E IS L GPACPEAAE +IG VL+ESLQYLQW++IG+FGG S SSW +VELW+ VE+EL+ HF ITA +RY++ LLOE L +SS+ARVIL CSSEDA
SGK050	57	114	partial	PK-like	GCyc	GCyc	1.30E-10	48	23	48	71	NP_003986.2	Natriuretic peptide receptor B precursor, isoform b (Homo sapiens)	17000035767558	R++ ++RF G CI P++ IV EYCP+GSL+D+LND++WF



Table 5 describes domains in the proteins outside of the kinase catalytic domain. The column headings are: "Gene Name", "ID#na", "ID#aa", "Extracatalytic Domains (AA boundaries)". Extracatalytic domains were identified by performing hidden Markov searches of the amino acid sequences using Pfam, a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains. Version 5.1 of Pfam (Sept 2000) contains alignments and models for 2015 protein families (<http://pfam.wustl.edu/faq.shtml>). The PFAM alignments were downloaded from <http://pfam.wustl.edu/hmmsearch.shtml> and the HMMr searches were run locally on a Timelogic computer (TimeLogic Corporation, Incline Village, NV). The PFAM accession number, the length in amino acids and the number of proteins used to build the profile are listed below.

The CNH domain (Pfam PF00780) is approximately 300 amino acids long. It is built from 23 members and found in NIK1-like kinase, mouse citron and yeast ROM1 and ROM2. The PKC terminal domain (PF00433) is approximately 66 amino acids long. It is built from 235 members and found in protein kinase C from multiple species. The phorbol esters/diacylglycerol-binding domain (C1 domain) (PF00130) is approximately 50 amino acids long. It is built from 269 members and in found in protein kinase C from multiple species. The RGS regulator of G protein signaling domain (PF00615) is approximately 125 amino acids long. It is built from 103 members and found in RGS (Regulator of G protein Signaling) family members that include the GTPase-activating proteins for heterotrimeric G-protein alpha-subunits. The PDZ domain (PF00595) is approximately 83 amino acids long. It is built from 721 members and found in membrane-associated proteins that include homologues of the MAGUK family of guanylate kinases, several protein phosphatases and protein kinases. PDZ domains are also found in neuronal nitric oxide synthase as well as in the subfamily of dystrophin-associated proteins, collectively known as syntrophins. The Octicosapeptide domain (PF00564) is approximately 30 amino acids long. It is built from 47 members that include NADPH oxidase subunits, sorting nexins and PtdIns 3-kinases. This motif may be involved in Ca<sup>++</sup> binding. The cyclin domain (PF00134) is approximately 267 amino acids long. It is built from 233 members that include cyclins, TFIIB and RB/p107. The RNA recognition domain (also known as RRM, RBD, or RNP) (PF00076) is approximately 71 amino acids long. It is built from 1335 members that

include a variety of RNA-binding proteins such hnRNP proteins, proteins implicated in regulation of alternative splicing, and protein components. The motif is also found in a few single-stranded DNA-binding proteins. The myosin head domain (PF00063) is approximately 409 amino acids long. It is built from 310 members that include the motor proteins such as myosin. The ankyrin domain (PF00023) is approximately 33 amino acids long. It is built from 2220 members that include the ankyrin family of structural proteins, CDK inhibitors such as p19INK4d, and other signaling proteins such as the nuclear factor NF-kappa-b p50 subunit and Bcl3 (b-cell lymphoma 3-encoded protein) among others. The ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure. The ephrin Receptor ligand binding domain (PF01404) is approximately 171 amino acids long. It is built from 52 members that include the Eph family of receptor tyrosine kinases. The fibronectin type III domain (PF00041) is approximately 85 amino acids long. It is built from 2468 members that include a variety of transmembrane and membrane-associated proteins that include fibronectin, cytokine receptors, receptor tyrosine kinases, receptor tyrosine phosphatases, etc. The SAM domain (Sterile alpha motif) (PF00536) is approximately 110 amino acids long. It is built from 64 members. The SAM domain is an evolutionarily conserved protein-binding region that is involved in the regulation of numerous developmental processes in diverse eukaryotes. The SAM domain can potentially function as a protein interaction module through its ability to homo- and hetero oligomerise with other SAM domains. The DAG Diacylglycerol (DAG) domain (PF00609) is approximately 166 amino acids long. It is built from 27 members from the diacylglycerol kinase subfamily of protein kinases. This domain is assumed to be an accessory domain in diacylglycerol binding. The ROI1 domain (PF01163) is approximately 570 amino acids in length and is known generally in the art (*see, e.g.* www.pfam.wustl.edu). It is built from 14 members and is believed to encode an atypical A-Kinase.

Table 5 - Extracatalytic Domains  
413406\_1.xls

Gene Name	ID#na	ID#aa	Extracatalytic Domains (AA boundaries)	Genomic_ID	Comparison	Cel Sort
SGK187, CRIK	1	58	CNH domain (1620-1917); PH 1472-1591; Protein Kinase C terminal domain (362-391); Phorbol esters/diacylglycerol binding domain (C1 domain) (1391-1439)	4926912_1_1		44
SGK064, GRK7	2	59	RGS (55-176)	11000284009826		23 4_SEQ_ID_1_11000284009826
SGK409, KIAA0303	3	60	PDZ domain (1020-1148)	6671897_1_1		54
SGK410	5	62	Protein kinase C terminal domain (558-624); Phorbol esters/diacylglycerol binding domain (C1 domain) (176-225); Octicosapeptide repeat (100-129)	7018620_2		55
SGK093, Wnk3	31	88	Cyclin (1373-1410)	11000284063441		33 4_SEQ_ID_38_11000284063441
SGK295, KIS	34	91	RNA recognition motif (345-401)	17000113079883		49
SGK125, MYO3A	36	93	Myosin head (340-1040); IQ, (3 domains, 1055-1366)	17000078090910		38
SGK127	38	95	Phorbol esters/diacylglycerol binding domain (C1 domain) (408-451)	17000062804843		39
SGK009, ANKRD3	39	96	Ankyrin (10 domains: 437-469, 470-502, 503-535, 536-568, 569-602, 602-635, 636-668, 669-701, 702-730, 738-770)	11000258161587	+DLGTTDVGKK KLVDIVSGDTS	6 1_SEQ_ID_37_11000258161587

Table 5 - Extracatalytic Domains  
413406\_1.xls

Gene Name	ID#na	ID#aa	Extracatalytic Domains (AA boundaries)	Genomic_ID	Comparison C+G+P+EV	Cell Sort
SGK188, EphA9	53	110	Ephrin receptor ligand binding domain (35-211); Fibronectin type III domain (2 domains: 339-436 & 454-537); SAM domain (Sterile alpha motif)(931-996)	17000057739161	LLDSK Q ELGW + PS+GWEEISGV DEH PIRTYQVCNV++ +Q+NWL+T W+ R Q++VEL+FTLR DC+SIP GTCKETFN+YY +E++ +R + KIDTIAADESFT Q DLG+R +KLNTE+RE+G P++++GF+LAFQ DVGACVALVSV RVY+K+C TV+ LA FP T + +LVEV G+CV +S+ E PPRM+C +GEWLP+G+C +C+AG	45
SGK390	55	112	Diacylglycerol kinase catalytic domain (332-457); DAG_PE-bind (176-225 & 248-298); PH domain (66-158) Diacylglycerol kinase accessory domain (770-927);	17000112562523		52

Table 6 shows the results of a gene expression analysis of the kinases presented in this application using a microarray of cDNAs derived from 469 tissues and cell lines. The cDNAs were spotted on nylon and probed with labeled kinase genes, as described in Materials and Methods below. The kinase probes were PCR cloned from genomic exons. Data presentation from left to right is as follows: "Tissue": tissue type of the cDNA; "Tumor sym", indicates that the tissue is derived from a tumor, "sym" refers to the fact that the 5' and 3' primers used to make the sample are the same; "Normal Sym", indicates normal tissue was used to make the sample, with symmetric primers as described above; "Tumor 1o", indicates that primary tumor tissue was used to make the cDNA; "Tumor cells", indicates that these cDNA samples were made from cultured tumor cells; "Normal", indicates that these samples are derived from normal tissue or cell lines; "Endos", indicates that these samples are derived from endothelium-related tissue sources; "p53" refers to the status, mutant or wild-type, of the p53 gene in the source samples. Normalized expression values are presented for each gene referred to by its SEQ\_ID# on the subsequent columns. Genes represented in Table 6 are: SGK187 (SEQ ID NO: 1); SGK124 (SEQ ID NO: 9); SGK386 (SEQ ID NO: 21); SGK003 (SEQ ID NO: 22); SGK093 (SEQ ID NO: 31); SGK074 (SEQ ID NO: 32); and SGK396 (SEQ ID NO: 43).

Table 6  
Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
TCGP	26							212	0	0	2,366	0	58,451	2,077
h keratinocytes 2/25/92 #10	46							83	3,050	0	2,340	0	89,541	1,223
h adult SMC 10/21/92 #17	47							0	0	0	1,713	374	73,933	291
h fibroblasts 3/31/92 #12	48							0	0	0	423	152	62,710	1,235
HOP-92	97							514	0	0	2,631	94	87,584	1,316
OVCAR-8	98							18	0	0	0	0	4,204	164
EKVX	99							19	0	0	231	0	27,721	105
IGROV1	100							0	0	0	65	0	7,800	0
NCI-H23	101							0	0	0	0	0	28,620	70
SK-OV-3	102							280	0	0	1,474	0	65,008	1,516
NCI-H226	103							153	365	0	1,091	91	107,932	3,834
SNB-19	104							0	0	0	855	0	74,412	1,302
NCI-H322M	105							243	77	0	422	226	35,551	84
SNB-75	106							138	0	0	0	0	98,053	1,787
NCI-H460	107							0	548	0	3,348	81	78,829	1,166
U251	108							0	554	0	0	0	84,310	677
NCI-H522	109							196	0	0	162	0	22,554	497
SF-268	110							0	771	0	0	593	49,287	1,115
AS49/ATCC	111							0	266	0	1,613	1,619	43,786	1,448
SF-295	112							0	0	0	1,333	0	7,106	0
HOP-62	113							231	343	0	529	1,746	21,968	670
SF-539	114							176	101	0	237	167	39,588	1,662
OVCAR-3	115							0	178	0	998	442	28,838	737
CCRF-CEM	116							80	6,816	0	605	0	45,975	2,130
OVCAR-4	117							52	57	0	704	292	34,439	1,510
K-562	118							0	0	0	0	185	33,123	1,072
OVCAR-5	119							0	0	0	1,307	542	18,184	23
MOLT-4	120							0	0	0	118	133	35,133	410
HL-60	121							0	537	0	0	658	80,247	719
SN12C	122							32	0	0	0	711	76,151	1,579
RPMI 8226	123							0	0	0	0	0	15,177	419
A498	124							0	904	0	430	3,358	122,082	982
SR	125							80	0	0	0	553	97,140	1,884
Caki-1	126							0	0	0	1,099	6,386	88,518	1,978
DU-145	127							560	191	0	169	0	3,687	0
RFX 393	128							51	202	0	685	1,192	108,097	1,825
PC-3	129							0	0	0	0	55	27,208	99
ACHN	130							0	0	0	1,248	0	21,998	0
HCC-2998	131							0	0	0	0	0	14,182	0
786-0	132							0	443	0	500	302	74,251	689
HCT 116	133							489	149	0	397	0	2,376	0
TK-10	134							94	102	0	1,682	415	51,674	1,842
SW-620	135							503	25	148	21,381	90	479,097	853
LOX IMVI	136							865	0	0	0	610	8,504	241
COLO 205	137							608	348	0	223	628	60,452	610
Malme-3M	138							134	260	0	416	734	33,878	1,065
HCT-15	139							69	281	0	0	781	35,146	562

[illegible]



**Table - Tissue Array  
413406\_1.xls**

[illegible]



Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 10	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
SK-MEL-2	140							720	219	0	1,523	340	21,933	848
KM-12	141							0	227	0	0	0	23,168	0
SK-MEL-5	142							346	1,100	0	1,862	319	20,650	825
UO-31	143							1,235	0	0	0	0	43,858	549
SK-MEL-28	144							0	121	0	362	151	2,283	158
UACC-62	145							0	74	0	473	1,045	106,158	1,533
UACC-257	147							1,027	131	0	1,889	0	132,416	2,031
M14	149							0	182	0	67	0	14,280	59
MCF7	151							86	0	108	17,887	0	478,093	791
MCF-7/ADR-RES	153							0	0	0	972	484	87,261	1,487
Hs 578T	155							0	0	0	1,717	0	125,561	611
MDA-MB-231	157							0	416	0	0	107	77,227	0
MDA-MB-435	159							391	359	0	1,576	0	41,654	1,257
MDA-N	161							0	183	0	2,141	380	28,371	1,457
T-47D	163							139	103	0	0	202	52,411	220
adrenal gland - h		1						64	29	0	16,403	133	808,208	1,624
lymph node - h		2						0	858	0	2,522	0	126,297	6,190
bone marrow - h		3						0	793	0	157	435	127,325	3,459
mammary gland - h		4						0	0	0	0	221	60,842	482
brain - h		5						0	309	0	206	884	109,918	4,188
pancreas - h		6						0	400	0	1,126	0	103,318	2,814
cerebellum - h		7						0	87	0	9,410	578	167,028	9,894
pituitary gland - h		8						83	1,273	0	1,820	194	108,893	5,097
fetal brain - h		9						359	1,004	0	4,928	882	121,604	8,651
placenta - h		10						0	0	0	6,212	1,310	115,812	2,852
fetal kidney - h		11						0	0	0	4,629	28,058	192,142	4,075
prostate, h		12						0	280	0	851	0	18,517	0
fetal liver - h		13						5,768	69	0	2,652	7,321	128,455	2,039
salivary gl. - h		14						0	0	0	1,389	0	92,748	1,033
fetal lung - h		15						0	666	0	3,837	1,462	198,037	4,592
skeletal muscle - h		16						1,336	0	0	2,895	0	43,019	42
heart - h		17						12,477	125	0	2,160	205	107,943	899
small intestine - h		18						0	226	0	1,736	1,221	117,752	862
kidney - h		19						0	56	0	2,418	18,839	135,359	1,267
spinal cord - h		20						0	144	0	1,031	451	119,848	789
liver - h		21						378	1,481	0	433	0	153,048	852
Spleen - h		22						778	191	0	861	0	122,519	1,248
lung - h		23						0	530	0	1,794	0	76,316	733
stomach - h		24						372	0	0	171	0	39,162	278
testis - h		25						15	0	0	14,003	0	141,552	5,293
thymus - h		27						1,943	343	0	2,504	0	180,985	6,498
HPAEC		28			28			0	0	0	2,192	0	171,007	2,973
thyroid gland - h		29						0	561	0	578	0	155,558	2,783
RPTC		30			30			166	170	0	2,416	0	40,468	0
trachea - h		31						0	5	0	2,843	618	126,079	3,124
HMEC		32						0	0	0	0	0	14,287	0
uterus - h		33						0	492	0	2,700	171	120,188	3,561

Table - Tissue Array  
413406\_1.xls

Tissue_type	Release	tag	pos	NCI/cac	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
SKMEL-2	R13841	sym	E20	269														
KMF-12	R13841	sym	F21	358														
SKMEL-5	R13841	sym	F21	358														
SKMEL-5	R13841	sym	F21	270														
UO-31	R13841	sym	F23	259														
SKMEL-28	R13841	sym	F24	371														
UACC-62	R13841	sym	G01	272														
UACC-257	R13841	sym	G01	273														
M14	R13841	sym	G05	374														
UACC-62	R13841	sym	G07	275														
UACC-257	R13841	sym	G09	276														
UACC-62	R13841	sym	G11	277														
UACC-257	R13841	sym	G13	278														
UACC-62	R13841	sym	G15	279														
UACC-257	R13841	sym	G17	280														
UACC-62	R13841	sym	G19	281														
UACC-257	R13841	sym	A01	285														
UACC-62	R13841	sym	A02	297														
UACC-257	R13841	sym	A03	286														
UACC-62	R13841	sym	A04	278														
UACC-257	R13841	sym	A05	287														
UACC-62	R13841	sym	A06	299														
UACC-257	R13841	sym	A07	288														
UACC-62	R13841	sym	A08	300														
UACC-257	R13841	sym	A09	289														
UACC-62	R13841	sym	A10	301														
UACC-257	R13841	sym	A11	290														
UACC-62	R13841	sym	A12	302														
UACC-257	R13841	sym	A13	291														
UACC-62	R13841	sym	A14	303														
UACC-257	R13841	sym	A15	292														
UACC-62	R13841	sym	A16	304														
UACC-257	R13841	sym	A17	293														
UACC-62	R13841	sym	A18	305														
UACC-257	R13841	sym	A19	294														
UACC-62	R13841	sym	A20	306														
UACC-257	R13841	sym	A21	295														
UACC-62	R13841	sym	A22	307														
UACC-257	R13841	sym	A23	296														
UACC-62	R13841	sym	A24	308														
UACC-257	R13841	sym	B01	309														
UACC-62	R13841	sym	B03	310														
UACC-257	R13841	sym	B04	319														
UACC-62	R13841	sym	B05	311														
UACC-257	R13841	sym	B06	310														
UACC-62	R13841	sym	B07	312														
UACC-257	R13841	sym	B08	321														
UACC-62	R13841	sym	B09	313														

**Table - Tissue Array**  
**413406\_1.xls**

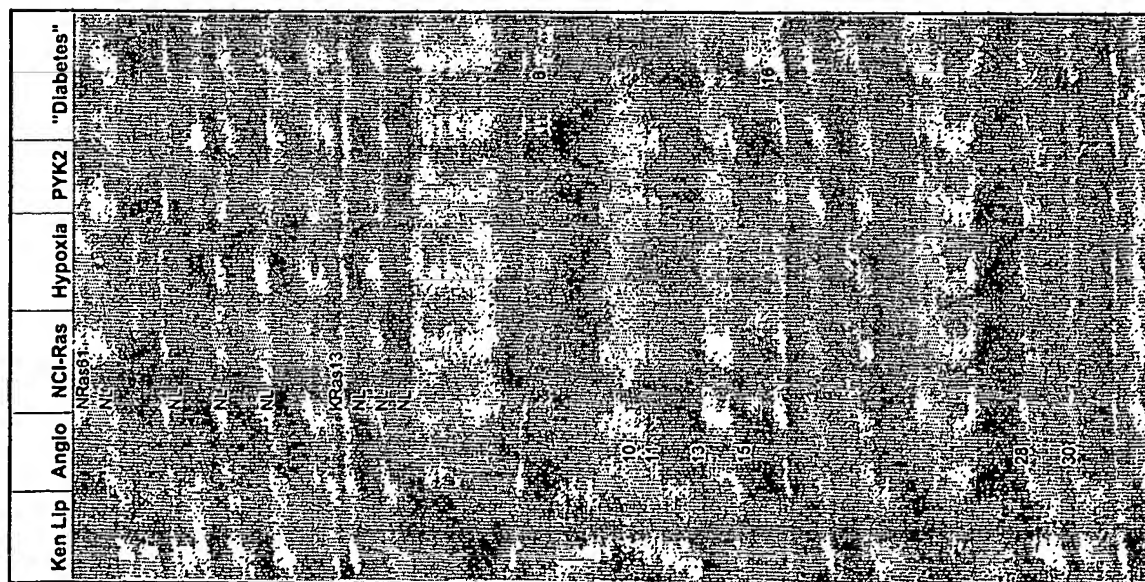


Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
HCAEC		34						495	0	0	0	300	11,188	0
Pancreas - h		35						0	87	0	482	0	0	329
lymph node - h		36						326	298	0	0	0	29,237	0
Skeletal muscle - h		37						242	0	0	1,216	0	16,454	0
fetal liver - h		38						0	0	0	506	0	164,981	412
Heart - h		39						500	94	0	2,052	0	135,384	2,681
thymus - h		40						0	166	0	145	44	84,824	394
Duodenum - h		41						717	113	0	2,188	421	98,060	2,415
Fetal brain - h		42						490	0	0	0	0	70,179	559
Salivary gl. - h		43						189	0	0	0	0	32,869	359
testis - h		44						0	221	0	396	0	45,754	196
fetal kidney - h					49			1,124	0	0	0	0	83,131	273
fetal liver - h					51			443	0	0	0	0	87,258	423
fetal lung - h					53			0	109	0	1,903	89	18,565	208
heart - h					55			0	45	0	268	0	32,265	251
Kidney - h					57			0	0	0	45	0	23,292	280
lung - h					59			0	183	0	0	0	84,812	1,129
lymph node - h					61			45	0	0	0	0	80,505	199
WI-38 72h					179			0	149	0	0	40	143,506	3,584
thymus - h					193			0	0	0	0	81	17,291	327
HEPM 3d TGFβ1 delergent+DNase					195			69	134	0	0	659	60,103	252
Salivary gl. - h					197			0	0	0	278	0	26,232	166
testis - h					199			0	86	0	0	17	11,196	143
Pancreas - h					201			199	0	0	98	0	32,630	0
Skeletal muscle - h					203			0	0	0	0	32	50,700	0
Duodenum - h					205			0	0	0	0	0	64,559	197
HMEC					209			777	102	0	0	206	33,654	0
fetal brain - h					210			0	69	0	1,050	0	101,094	3,485
HCAEC					211			0	210	0	289	313	26,545	248
placenta - h					212		211	0	296	0	0	0	106,354	2,700
fetal liver - h					213			1,358	40	0	0	0	16,093	223
stomach - h					214			0	0	0	151	0	47,791	176
Heart - h					215			0	0	0	92	0	26,452	196
bladder - h					222			0	140	0	0	0	51,494	0
HT372-normal					234			0	50	0	0	0	39,930	0
HT382-normal					268			615	78	0	0	0	19,299	10
HT392-normal					268			131	0	0	0	0	86,365	202
HPAEC					275		275	0	0	0	563	286	30,692	361
adrenal gland - h					277			2,553	446	166	27,369	329	1,532,778	4,110
bone marrow - h					279			0	19	0	0	0	6,744	114
skeletal muscle - h					280			0	45	0	1,545	0	102,189	1,579
small intestine - h					282			0	0	0	0	0	40,631	318
spinal cord - h					284			0	0	0	357	0	48,103	502
Spleen - h					296			0	3	0	657	153	71,535	60
liver - h					297			375	0	0	0	0	88,510	74
testis - h					298			422	0	0	640	0	167,195	2,423
bladder - h					302			38	311	0	2,206	0	74,917	825



Table - Tissue Array  
413406\_1.xls

Tissue type	Release	tag	pos	NCI/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
HCAEC	R1384-1	5ym	B10	322														
Pancreas	R1384-1	5ym	B11	314														
Lymph node	R1384-1	5ym	B12	323														
Skeletal muscle	R1384-1	5ym	B13	315														
Fetal liver	R1384-1	5ym	B14	324														
Heart	R1384-1	5ym	B15	316														
Thymus	R1384-1	5ym	B16	325														
Duodenum	R1384-1	5ym	B17	317														
Salivary gland	R1384-1	5ym	B18	326														
Testis	R1384-1	5ym	B19	318														
Fetal kidney	R1384-1	5ym	B20	327														
Fetal liver	R1384-1	5ym	C01	321														
Testis	R1384-1	5ym	C02	328														
Heart	R1384-1	5ym	C03	329														
Brain	R1384-1	5ym	C04	330														
Prostate	R1384-1	5ym	C05	331														
Adipose	R1384-1	5ym	C06	332														
Lung	R1384-1	5ym	C07	333														
Myoblast	R1384-1	5ym	C08	334														
Myoblast	R1384-1	5ym	C09	335														
Myoblast	R1384-1	5ym	C10	336														
Myoblast	R1384-1	5ym	C11	337														
Myoblast	R1384-1	5ym	C12	338														
Myoblast	R1384-1	5ym	C13	339														
Myoblast	R1384-1	5ym	C14	340														
Myoblast	R1384-1	5ym	C15	341														
Myoblast	R1384-1	5ym	C16	342														
Myoblast	R1384-1	5ym	C17	343														
Myoblast	R1384-1	5ym	C18	344														
Myoblast	R1384-1	5ym	C19	345														
Myoblast	R1384-1	5ym	C20	346														
Myoblast	R1384-1	5ym	C21	347														
Myoblast	R1384-1	5ym	C22	348														
Myoblast	R1384-1	5ym	C23	349														
Myoblast	R1384-1	5ym	C24	350														
Myoblast	R1384-1	5ym	C25	351														
Myoblast	R1384-1	5ym	C26	352														
Myoblast	R1384-1	5ym	C27	353														
Myoblast	R1384-1	5ym	C28	354														
Myoblast	R1384-1	5ym	C29	355														
Myoblast	R1384-1	5ym	C30	356														
Myoblast	R1384-1	5ym	C31	357														
Myoblast	R1384-1	5ym	C32	358														
Myoblast	R1384-1	5ym	C33	359														
Myoblast	R1384-1	5ym	C34	360														
Myoblast	R1384-1	5ym	C35	361														
Myoblast	R1384-1	5ym	C36	362														
Myoblast	R1384-1	5ym	C37	363														
Myoblast	R1384-1	5ym	C38	364														
Myoblast	R1384-1	5ym	C39	365														
Myoblast	R1384-1	5ym	C40	366														
Myoblast	R1384-1	5ym	C41	367														
Myoblast	R1384-1	5ym	C42	368														
Myoblast	R1384-1	5ym	C43	369														
Myoblast	R1384-1	5ym	C44	370														
Myoblast	R1384-1	5ym	C45	371														
Myoblast	R1384-1	5ym	C46	372														
Myoblast	R1384-1	5ym	C47	373														
Myoblast	R1384-1	5ym	C48	374														
Myoblast	R1384-1	5ym	C49	375														
Myoblast	R1384-1	5ym	C50	376														
Myoblast	R1384-1	5ym	C51	377														
Myoblast	R1384-1	5ym	C52	378														
Myoblast	R1384-1	5ym	C53	379														
Myoblast	R1384-1	5ym	C54	380														
Myoblast	R1384-1	5ym	C55	381														
Myoblast	R1384-1	5ym	C56	382														
Myoblast	R1384-1	5ym	C57	383														
Myoblast	R1384-1	5ym	C58	384														
Myoblast	R1384-1	5ym	C59	385														
Myoblast	R1384-1	5ym	C60	386														
Myoblast	R1384-1	5ym	C61	387														
Myoblast	R1384-1	5ym	C62	388														
Myoblast	R1384-1	5ym	C63	389														
Myoblast	R1384-1	5ym	C64	390														
Myoblast	R1384-1	5ym	C65	391														
Myoblast	R1384-1	5ym	C66	392														
Myoblast	R1384-1	5ym	C67	393														
Myoblast	R1384-1	5ym	C68	394														
Myoblast	R1384-1	5ym	C69	395														
Myoblast	R1384-1	5ym	C70	396														
Myoblast	R1384-1	5ym	C71	397														
Myoblast	R1384-1	5ym	C72	398														
Myoblast	R1384-1	5ym	C73	399														
Myoblast	R1384-1	5ym	C74	400														
Myoblast	R1384-1	5ym	C75	401														
Myoblast	R1384-1	5ym	C76	402														
Myoblast	R1384-1	5ym	C77	403														
Myoblast	R1384-1	5ym	C78	404														
Myoblast	R1384-1	5ym	C79	405														
Myoblast	R1384-1	5ym	C80	406														
Myoblast	R1384-1	5ym	C81	407														
Myoblast	R1384-1	5ym	C82	408														
Myoblast	R1384-1	5ym	C83	409														
Myoblast	R1384-1	5ym	C84	410														
Myoblast	R1384-1	5ym	C85	411														
Myoblast	R1384-1	5ym	C86	412														
Myoblast	R1384-1	5ym	C87	413														
Myoblast	R1384-1	5ym	C88	414														
Myoblast	R1384-1	5ym	C89	415														
Myoblast	R1384-1	5ym	C90	416														
Myoblast	R1384-1	5ym	C91	417														
Myoblast	R1384-1	5ym	C92	418														
Myoblast	R1384-1	5ym	C93	419														
Myoblast	R1384-1	5ym	C94	420														
Myoblast	R1384-1	5ym	C95	421														
Myoblast	R1384-1	5ym	C96	422														
Myoblast	R1384-1	5ym	C97	423														
Myoblast	R1384-1	5ym	C98	424														
Myoblast	R1384-1	5ym	C99	425														
Myoblast	R1384-1	5ym	C100	426														
Myoblast	R1384-1	5ym	C101	427														
Myoblast	R1384-1	5ym	C102	428														
Myoblast	R1384-1	5ym	C103	429														
Myoblast	R1384-1	5ym	C104	430														
Myoblast	R1384-1	5ym	C105	431														
Myoblast	R1384-1	5ym	C106	432														
Myoblast	R1384-1	5ym	C107	433														
Myoblast	R1384-1	5ym	C108	434														

**Table - Tissue Array**  
**413406\_1.xls**

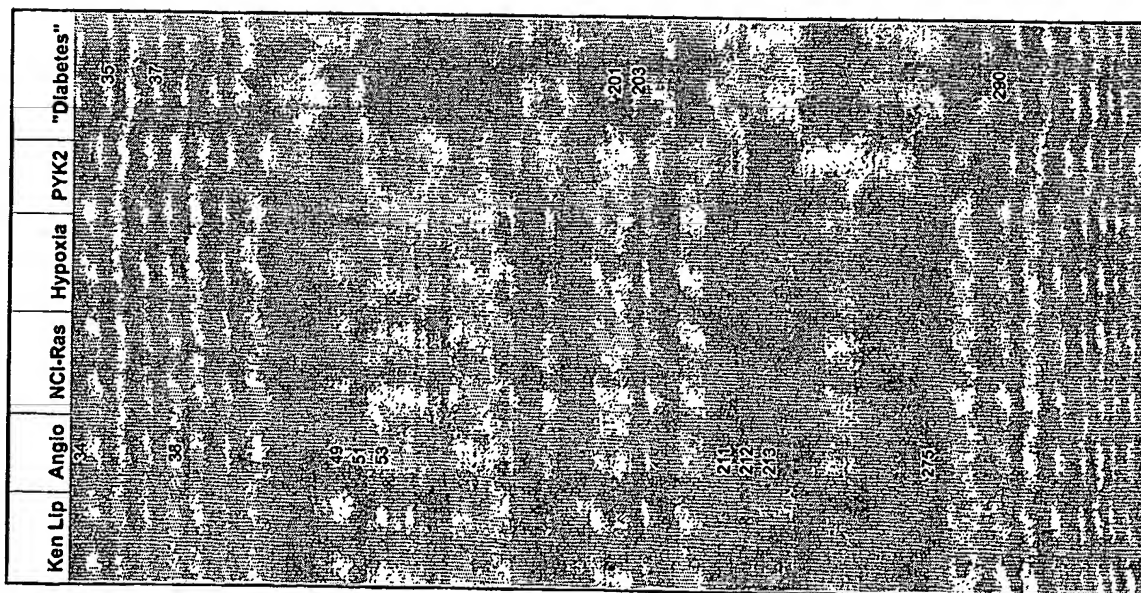


Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 10	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
mammary gland - h					303			221		0	794	117	48,744	0
pancreas - h					305			678	16	0	817	473	127,476	342
pituitary gland - h					307			0	0	0	0	0	20,657	0
prostate, h					309			0	634	0	297	2,243	148,345	3,258
salivary gl. - h					311			455	0	0	1,109	0	18,958	0
thyroid gland - h					314			0	0	0	0	0	127,246	72
trachea - h					316			0	125	0	16	172	80,358	0
uterus - h					318			19	0	0	1,641	0	73,044	1,084
HEPM 3d untreated					320			1,085	18	0	0	510	61,981	0
HT149 - normal					321			528	0	0	0	0	56,126	155
thymus, h					326			0	0	0	0	0	51,419	0
HT398-normal					327			551	431	0	742	0	32,502	24
facial brain - h					328			734	0	0	1,133	0	166,205	3,822
h adult SMC 10/21/92 #17					330			0	0	0	0	0	23,012	164
lymph node - h					332			95	0	0	0	67	22,356	0
APTEC					334	334		1,186	189	0	223	0	96,951	2,705
brain - h					342			0	204	0	0	716	51,925	0
cerebellum - h					344			0	0	0	40	0	31,053	0
HT157-normal					361			0	285	0	0	0	81,782	1
HT213-normal					363			1,947	45	0	15	0	20,122	0
HT218-normal					365			0	0	0	0	0	13,192	0
HeLa - 1							HPV E6	188	27	0	17,506	191	555,591	1,745
HeLa - 2							HPV E6	362	758	2	52,193	0	679,609	1,613
HeLa - 3							HPV E6	1,051	0	67	19,138	0	443,018	1,179
HeLa - 4							HPV E6	699	387	0	11,837	0	636,546	1,133
HeLa - 5							HPV E6	1,571	0	600	40,632	0	740,204	1,934
HeLa - 6							HPV E6	0	0	215	40,279	174	849,472	3,502
HeLa - 7							HPV E6	233	0	183	39,006	81	914,878	2,315
HeLa - 8							HPV E6	0	0	0	30,268	0	967,084	1,982
ADR-RES - 1							mutant	540	188	0	48,622	0	921,921	922
ADR-RES - 2							mutant	76	0	0	18,959	0	1,010,246	591
ADR-RES - 3							mutant	334	344	0	13,658	0	359,976	383
ADR-RES - 4							mutant	0	0	0	8,302	0	442,971	11
ADR-RES - 5							mutant	306	0	164	20,910	0	544,895	130
ADR-RES - 6							mutant	140	0	0	26,446	0	601,648	468
ADR-RES - 7							mutant	0	681	0	1,745	429	85,849	3,468
ADR-RES - 8							mutant	0	283	0	3,363	0	5,866	0
C33A - 1							mutant	0	0	0	3,630	0	458,295	578
C33A - 2							mutant	0	0	0	13,815	0	389,598	609
C33A - 3							mutant	4,866	0	104	9,604	0	379,820	824
C33A - 4							mutant	89	0	128	8,326	42	498,796	1,112
C33A - 5							mutant	0	0	0	34,820	0	644,383	890
C33A - 6							mutant	0	0	0	680	0	678,800	0
C33A - 7							mutant	0	0	149	18,136	0	785,878	908
C33A - 8							mutant	21	0	173	42,512	0	1,172,556	2,189
EKVX - 1							mutant	675	0	0	27,186	0	269,251	221
EKVX - 2							mutant	0	0	92	69,264	272	498,108	0



Table - Tissue Array  
413406\_1.xls

Tissue_type	Release	tag	pos	NC/etc	CC	DD	MC	SR	ANG	PRO	MetS	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
mammary gland - h	R1384-1	asym	M13	100														
Pancreas - h	R1384-1	asym	M17	100														
adipose - h	R1384-1	asym	M18	100														
salivary gland - h	R1384-1	asym	M19	102														
prostate - h	R1384-1	asym	M21	104								309						
SA-OS (Mundy) poly A+	R1384-1	asym	M23	105														
thyroid gland - h	R1384-1	asym	N02	113														
trachea - h	R1384-1	asym	N04	114														
uterus - h	R1384-1	asym	N06	115														
HEPIM3d untreated	R1384-1	asym	N08	171														
HT149 normal	R1384-1	asym	N09	163														
thymus - h	R1384-1	asym	N14	127										326				
HT398 normal	R1384-1	asym	N15	219														
Fetal brain - h	R1384-1	asym	N16	128									328					
head SMC 102762 #17	R1384-1	asym	N18	75														
lymph node - h	R1384-1	asym	N20	125														
RP1EC	R1384-1	asym	N22	122														
brain - h	R1384-1	asym	O06	89														
carabellum - h	R1384-1	asym	O08	90									342					
HT157 normal	R1384-1	asym	P01	168									344					
HT213 normal	R1384-1	asym	P03	181														
HT218 normal	R1384-1	asym	P05	183														
HeLa-1	R2384-1	sym	G10	394	CC-26		MC-31	SR-21						58	58	58	58	58
HeLa-2	R2384-1	sym	G18	395				SR-22						138	138	138	138	138
HeLa-3	R2384-1	sym	G12	396	CC-27													
HeLa-4	R2384-1	sym	G14	397	CC-28													
HeLa-5	R2384-1	sym	G16	398	CC-29													
HeLa-6	R2384-1	sym	G18	399														
HeLa-7	R2384-1	sym	G20	400	CC-30		MC-32											
HeLa-8	R2384-1	sym	G22	401			MC-33											
HeLa-9	R2384-1	sym	G17	347	CC-16	DD-28	MC-28	SR-19										
ADRES-3	R2384-1	sym	G19	342				SR-20										
ADRES-4	R2384-1	sym	G21	343	CC-17													
ADRES-5	R2384-1	sym	G23	344	CC-18	DD-29												
ADRES-6	R2384-1	sym	G02	345	CC-19													
ADRES-7	R2384-1	sym	G04	346		DD-30												
ADRES-8	R2384-1	sym	G06	347	CC-20		MC-29											
adrenal gland - h	R2384-1	sym	G08	348		MC-30												
adrenal gland - h	R2384-1	sym	D08	349		MC-40	SR-27								80	80	80	80
C33A-1	R2384-1	sym	D10	350			SR-28								82	82	82	82
C33A-2	R2384-1	sym	D12	351														
C33A-3	R2384-1	sym	D14	352														
C33A-4	R2384-1	sym	D16	353														
C33A-5	R2384-1	sym	D18	354														
C33A-6	R2384-1	sym	D20	355			MC-41											
C33A-7	R2384-1	sym	D22	356			MC-42											
C33A-8	R2384-1	sym	F02	370	CC-09	DD-04	MC-04	SR-03							122	122	122	122
EVX-2	R2384-1	sym	F04	371			SR-04								124	124	124	124



Table - Tissue Array  
413406\_1.xls

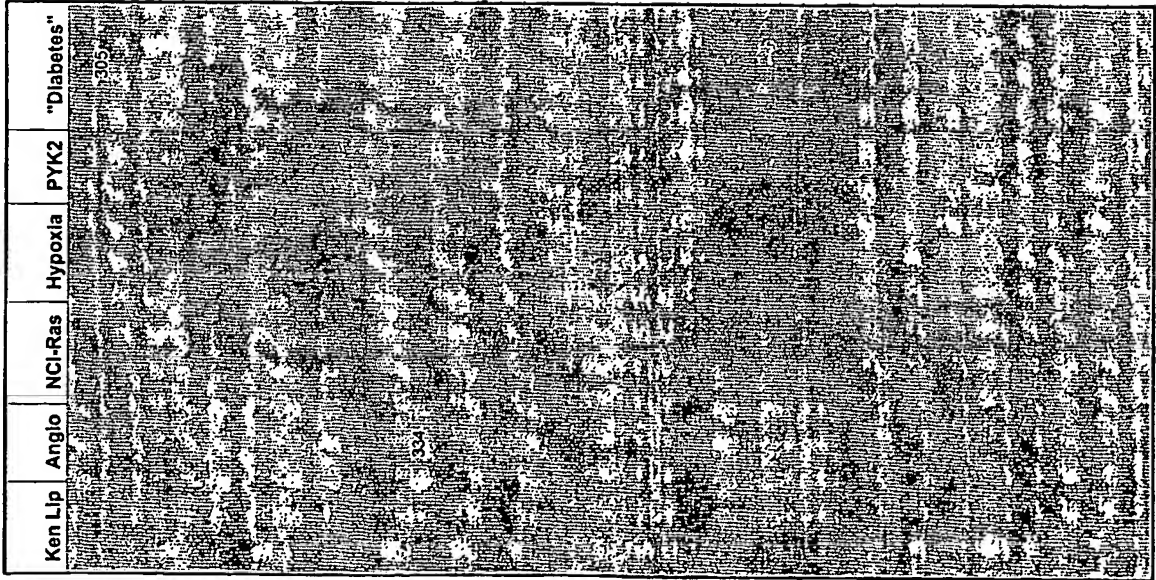


Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
EKVX - 3							mutant	0	4	0	93,678	0	429,070	340
EKVX - 4							mutant	577	0	71	140,696	344	429,429	819
EKVX - 5							mutant	28	0	0	34,966	0	813,745	1,003
EKVX - 6							mutant	6	0	98	7,689	79	459,086	799
EKVX - 7							mutant	705	0	564	41,327	393	1,353,897	2,803
EKVX - 8							mutant	192	242	0	31,342	0	397,494	1,472
H1299 - 1							mutant	1,259	0	62	26,353	0	501,247	1,055
H1299 - 2							mutant	0	327	37	13,882	0	712,338	1,004
H1299 - 3							mutant	0	121	16	36,682	0	778,091	2,440
H1299 - 4							mutant	172	0	22	57,330	669	957,276	2,317
H1299 - 5							mutant	0	0	92	18,628	0	865,896	1,039
H1299 - 6							mutant	0	79	52	17,264	172	338,356	412
H1299 - 7							mutant	147	0	0	12,801	0	523,628	0
H1299 - 8							mutant	2,032	0	0	21,626	553	519,154	449
H1299 - 1							mutant	817	0	85	60,106	0	859,216	1,413
H1299 - 2							mutant	1,145	298	16	62,721	0	440,697	1,568
H1299 - 3							mutant	833	228	132	85,651	228	700,898	958
H1299 - 4							mutant	120	0	258	102,716	604	1,046,464	1,512
H1299 - 5							mutant	810	0	0	137,429	0	1,522,264	3,278
H1299 - 6							mutant	3,401	128	154	131,385	568	1,157,087	2,992
H1299 - 7							mutant	0	0	25	21,556	455	359,692	523
H1299 - 8							mutant	0	0	227	34,495	11	539,304	1,099
OVCAR-5-1							mutant	166	0	0	46,515	403	407,563	1,031
OVCAR-5-2							mutant	0	0	81	22,899	15	550,448	653
OVCAR-5-3							mutant	66	344	186	67,219	0	386,258	33
OVCAR-5-4							mutant	0	81	0	28,479	1	460,438	2,500
OVCAR-5-5							mutant	16	314	0	52,491	140	607,764	883
OVCAR-5-6							mutant	0	144	41	5,128	0	632,947	91
OVCAR-5-7							mutant	1,388	743	379	73,397	0	2,126,515	1,252
OVCAR-5-8							mutant	687	0	0	72,830	1,009	411,234	1,283
SF-268-1							mutant	454	0	0	46,444	169	1,247,938	3,199
SF-268-2							mutant	175	0	1	31,761	0	294,676	519
SF-268-3							mutant	35	0	211	26,186	0	442,870	212
SF-268-4							mutant	178	184	204	51,244	0	399,799	568
SF-268-5							mutant	0	816	0	58,103	0	471,620	0
SF-268-6							mutant	0	38	0	25,570	0	394,514	619
SF-268-7							mutant	850	143	0	40,201	429	347,362	1,637
SF-268-8							mutant	2,391	138	180	67,712	12	463,334	1,254
SW480 - 1							mutant	0	0	0	26	0	13,937	30
SW480 - 2							mutant	470	255	0	1,000	79	33,878	1,032
SW480 - 3							mutant	79	0	0	76,080	0	1,460,381	3,005
SW480 - 4							mutant	260	85	206	45,130	226	423,651	537
SW480 - 5							mutant	733	0	0	33,127	0	392,824	235
SW480 - 6							mutant	1,608	0	0	11,557	0	536,855	116
SW480 - 7							mutant	183	0	0	38,957	0	729,185	983
SW480 - 8							mutant	1,263	125	0	27,056	0	555,267	1,473
U2OS - 1							mutant	0	0	0	78,276	1,905	1,630,846	4,317

Table - Tissue Array  
413406\_1.xls

Tissue_type	Release	tag	pos	NCI/letc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
EKVA-3	R2384	sym	E06	372	CC-08												128	
EKVA-4	R2384	sym	F03	373	CC-07	DD-05											128	
EKVA-5	R2384	sym	A17	374	CC-06												17	
EKVA-6	R2384	sym	A19	375		DD-06											21	19
EKVA-7	R2384	sym	A21	376	CC-10		MC-05										21	
EKVA-8	R2384	sym	F10	377			MC-06										130	
H1289-1	R2384	sym	D15	378	CC-31		MC-37	SR-25									87	87
H1289-2	R2384	sym	D17	379				SR-26									87	
H1289-3	R2384	sym	D19	380	CC-32												91	
H1289-4	R2384	sym	D21	381	CC-33												93	83
H1289-5	R2384	sym	D23	382	CC-34												85	
H1289-6	R2384	sym	D01	383													74	
H1289-7	R2384	sym	D04	384	CC-35		MC-38										76	
H1289-8	R2384	sym	D06	385			MC-39										78	
H1289-9	R2384	sym	A14	409		DD-10	MC-10										14	14
H1289-10	R2384	sym	A16	410				SR-03									18	
H1289-11	R2384	sym	A18	411													20	20
H1289-12	R2384	sym	A20	412		DD-11											22	
H1289-13	R2384	sym	A22	413													25	
H1289-14	R2384	sym	A24	414		DD-12											25	
H1289-15	R2384	sym	B01	415			MC-11										27	
H1289-16	R2384	sym	B03	416			MC-12										35	
H1289-17	R2384	sym	B17	454		DD-22	MC-22	SR-15									38	
H1289-18	R2384	sym	B14	455				SR-18									40	
H1289-19	R2384	sym	B16	456													42	
H1289-20	R2384	sym	B18	457		DD-23											42	
H1289-21	R2384	sym	H03	458													171	
H1289-22	R2384	sym	B22	459		DD-24											171	
H1289-23	R2384	sym	B24	460			MC-23										171	
H1289-24	R2384	sym	F16	461			MC-24										171	
H1289-25	R2384	sym	E24	462		DD-16	MC-16	SR-11					120				120	
H1289-26	R2384	sym	F01	463				SR-12					121				120	
H1289-27	R2384	sym	F03	464									123				123	
H1289-28	R2384	sym	F05	465		DD-17							125				125	
H1289-29	R2384	sym	F07	466									127				127	
H1289-30	R2384	sym	T09	467		DD-18							129				129	
H1289-31	R2384	sym	F19	468			MC-17						131				131	
H1289-32	R2384	sym	F15	469			MC-18						133				133	
H1289-33	R2384	sym	G24	470			MC-21	SR-23					135				135	
H1289-34	R2384	sym	D01	471				SR-24					137				137	
H1289-35	R2384	sym	D03	472									139				139	
H1289-36	R2384	sym	D05	473									141				141	
H1289-37	R2384	sym	D07	474									143				143	
H1289-38	R2384	sym	D09	475									145				145	
H1289-39	R2384	sym	D11	476			MC-35						147				147	
H1289-40	R2384	sym	D13	477			MC-36						149				149	
H1289-41	R2384	sym	D15	478			MC-37						151				151	
H1289-42	R2384	sym	D17	479			MC-38						153				153	
H1289-43	R2384	sym	D19	480			MC-39						155				155	
H1289-44	R2384	sym	D21	481			MC-40						157				157	
H1289-45	R2384	sym	D23	482			MC-41						159				159	
H1289-46	R2384	sym	D25	483			MC-42						161				161	
H1289-47	R2384	sym	D27	484			MC-43						163				163	
H1289-48	R2384	sym	D29	485			MC-44						165				165	
H1289-49	R2384	sym	D31	486			MC-45						167				167	
H1289-50	R2384	sym	D33	487			MC-46						169				169	
H1289-51	R2384	sym	D35	488			MC-47						171				171	
H1289-52	R2384	sym	D37	489			MC-48						173				173	
H1289-53	R2384	sym	D39	490			MC-49						175				175	
H1289-54	R2384	sym	D41	491			MC-50						177				177	
H1289-55	R2384	sym	D43	492			MC-51						179				179	
H1289-56	R2384	sym	D45	493			MC-52						181				181	
H1289-57	R2384	sym	D47	494			MC-53						183				183	
H1289-58	R2384	sym	D49	495			MC-54						185				185	
H1289-59	R2384	sym	D51	496			MC-55						187				187	
H1289-60	R2384	sym	D53	497			MC-56						189				189	
H1289-61	R2384	sym	D55	498			MC-57						191				191	
H1289-62	R2384	sym	D57	499			MC-58						193				193	
H1289-63	R2384	sym	D59	500			MC-59						195				195	
H1289-64	R2384	sym	D61	501			MC-60						197				197	
H1289-65	R2384	sym	D63	502			MC-61						199				199	
H1289-66	R2384	sym	D65	503			MC-62						201				201	
H1289-67	R2384	sym	D67	504			MC-63						203				203	
H1289-68	R2384	sym	D69	505			MC-64						205				205	
H1289-69	R2384	sym	D71	506			MC-65						207				207	
H1289-70	R2384	sym	D73	507			MC-66						209				209	
H1289-71	R2384	sym	D75	508			MC-67						211				211	
H1289-72	R2384	sym	D77	509			MC-68						213				213	
H1289-73	R2384	sym	D79	510			MC-69						215				215	
H1289-74	R2384	sym	D81	511			MC-70						217				217	
H1289-75	R2384	sym	D83	512			MC-71						219				219	
H1289-76	R2384	sym	D85	513			MC-72						221				221	
H1289-77	R2384	sym	D87	514			MC-73						223				223	
H1289-78	R2384	sym	D89	515			MC-74						225				225	
H1289-79	R2384	sym	D91	516			MC-75						227				227	
H1289-80	R2384	sym	D93	517			MC-76						229				229	
H1289-81	R2384	sym	D95	518			MC-77						231				231	
H1289-82	R2384	sym	D97	519			MC-78						233				233	
H1289-83	R2384	sym	D99	520			MC-79						235				235	
H1289-84	R2384	sym	D101	521			MC-80						237				237	
H1289-85	R2384	sym	D103	522			MC-81						239				239	
H1289-86	R2384	sym	D105	523			MC-82						241				241	
H1289-87	R2384	sym	D107	524			MC-83						243				243	
H1289-88	R2384	sym	D109	525			MC-84						245				245	
H1289-89	R2384	sym	D111	526			MC-85						247				247	
H1289-90	R2384	sym	D113	527			MC-86						249				249	
H1289-91	R2384	sym	D115	528			MC-87						251				251	
H1289-92	R2384	sym	D117	529			MC-88						253				253	
H1289-93	R2384	sym	D119	530			MC-89						255				255	
H1289-94	R2384	sym	D121	531			MC-90						257				257	
H1289-95	R2384	sym	D123	532			MC-91						259				259	
H1289-96	R2384	sym	D125	533			MC-92						261				261	
H1289-97	R2384	sym	D127	534			MC-93						263				263	
H1289-98	R2384	sym	D129	535			MC-94						265				265	
H1289-99	R2384	sym	D131	536			MC-95						267				267	
H1289-100	R2384	sym	D133	537														

**Table - Tissue Array  
413406\_1.xls**





























































Ken Lip	Anglo	NCI-Ras	Hypoxia	PYK2	"Diabetes"
					
					
					
					
					
					
					
					
					
					

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 10	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
U2OS - 2							mutant	641	327	202	29,915	0	432,337	972
U2OS - 3							mutant	573	0	1	40,362	127	408,481	412
U2OS - 4							mutant	1,545	0	162	10,515	106	399,039	738
U2OS - 5							mutant	679	0	94	37,160	0	491,594	827
U2OS - 6							mutant	0	294	108	5,845	3,882	371,040	1,175
U2OS - 7							mutant	0	0	0	22,921	17	500,363	1,551
U2OS - 8							mutant	0	0	0	16,035	0	472,901	1,979
A549 - 1							wt	0	0	112	54,563	0	546,984	523
A549 - 2							wt	1,553	0	0	25,728	182	490,791	367
A549 - 3							wt	0	0	0	8,583	0	526,257	173
A549 - 4							wt	56	118	92	27,986	170	637,978	439
A549 - 5							wt	0	119	0	58,834	0	470,171	657
A549 - 6							wt	0	0	203	2,773	0	560,302	671
A549 - 7							wt	240	0	188	32,976	0	658,479	1,455
A549 - 8							wt	57	157	1	74,102	243	408,754	1,458
HCT-116 - 1							wt	0	0	0	100,001	85	2,916,547	3,159
HCT-116 - 2							wt	0	18	4	145,047	0	386,158	1,870
HCT-116 - 3							wt	665	0	0	17,785	159	427,289	313
HCT-116 - 4							wt	275	0	0	38,703	0	386,481	198
HCT-116 - 5							wt	558	179	38	37,110	85	505,899	955
HCT-116 - 6							wt	754	0	105	61,500	0	498,369	488
HCT-116 - 7							wt	190	187	47	42,722	0	412,159	1,068
HCT-116 - 8							wt	783	0	49	48,607	133	555,183	1,382
Hs58 - 1							wt	0	0	50	21,139	0	562,011	1,788
Hs58 - 2							wt	21	0	369	52,149	175	832,186	1,856
Hs58 - 3							wt	1,647	0	0	52,724	119	1,087,919	3,110
Hs58 - 4							wt	0	0	0	104,514	40	271,530	521
Hs58 - 5							wt	0	0	47	5,341	425	491,608	246
Hs58 - 6							wt	367	0	0	25,147	0	498,513	192
Hs58 - 7							wt	0	16	0	30,704	0	604,049	344
MCF-7 - 1							wt	0	0	0	7,662	0	662,630	341
MCF-7 - 2							wt	1,092	0	141	26,075	0	632,258	621
MCF-7 - 3							wt	1,349	0	341	17,550	0	661,141	1,174
MCF-7 - 4							wt	275	6	0	20,701	0	484,087	737
MCF-7 - 5							wt	0	0	0	8,565	237	795,802	1,446
MCF-7 - 6							wt	0	0	0	20,587	386	33,599	0
MCF-7 - 7							wt	0	0	0	0	357	76,817	535
MCF-7 - 8							wt	560	0	0	2,415	460	777,026	2,150
OVCAR-4 - 1							wt	1,159	0	126	65,996	43	3,789,726	5,436
OVCAR-4 - 2							wt	0	478	0	48,716	0	417,174	388
OVCAR-4 - 3							wt	0	314	77	29,888	0	477,283	117
OVCAR-4 - 4							wt	0	118	0	10,373	0	581,094	506
OVCAR-4 - 5							wt	396	123	89	38,345	0	600,108	457
OVCAR-4 - 6							wt	462	56	0	41,745	0	427,310	554
OVCAR-4 - 7							wt	59	0	208	7,508	356	471,191	924
OVCAR-4 - 8							wt	278	0	121	88,152	0	509,238	91
SF539 - 1							wt	0	283	92	11,389	0		



Tissue type	Release	tag	pos	NC/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
U20S-2	R2-384	sym	E01	487				SR-30							97			
U20S-3	R2-384	sym	E01	488														
U20S-4	R2-384	sym	E01	489														
U20S-5	R2-384	sym	E05	489														
U20S-6	R2-384	sym	E07	490														
U20S-7	R2-384	sym	E09	491														
U20S-8	R2-384	sym	E11	492														
U20S-9	R2-384	sym	E13	493														
A5-9-1	R2-384	sym	A01	333 CC-01	DD-01	MC-01	SR-01											
A5-9-2	R2-384	sym	A03	334			SR-02											
A5-9-3	R2-384	sym	A05	335 CC-02														
A5-9-4	R2-384	sym	A07	336 CC-03	DD-02													
A5-9-5	R2-384	sym	A09	337 CC-04														
A5-9-6	R2-384	sym	A11	338	DD-03													
A5-9-7	R2-384	sym	A13	339 CC-05			MC-02											
A5-9-8	R2-384	sym	A15	340			MC-03											
A5-9-9	R2-384	sym	A21	386														
HGT-116-1	R2-384	sym	F12	387			DD-07 MC-07	SR-05										
HGT-116-2	R2-384	sym	A02	388				SR-06										
HGT-116-3	R2-384	sym	A04	389			DB-08											
HGT-116-4	R2-384	sym	A06	390														
HGT-116-5	R2-384	sym	A08	391			DD-08											
HGT-116-6	R2-384	sym	A10	392														
HGT-116-7	R2-384	sym	A12	393														
HGT-116-8	R2-384	sym	E15	402				SR-31										
H568-1	R2-384	sym	E17	403				SR-32										
H568-2	R2-384	sym	E19	404														
H568-3	R2-384	sym	E21	405														
H568-4	R2-384	sym	E23	406														
H568-5	R2-384	sym	E25	407														
H568-6	R2-384	sym	E27	408														
H568-7	R2-384	sym	E29	409														
H568-8	R2-384	sym	E31	410														
H568-9	R2-384	sym	E33	411														
H568-10	R2-384	sym	E35	412														
H568-11	R2-384	sym	E37	413														
H568-12	R2-384	sym	E39	414														
H568-13	R2-384	sym	E41	415														

Table - Tissue Array  
413406\_1.xls

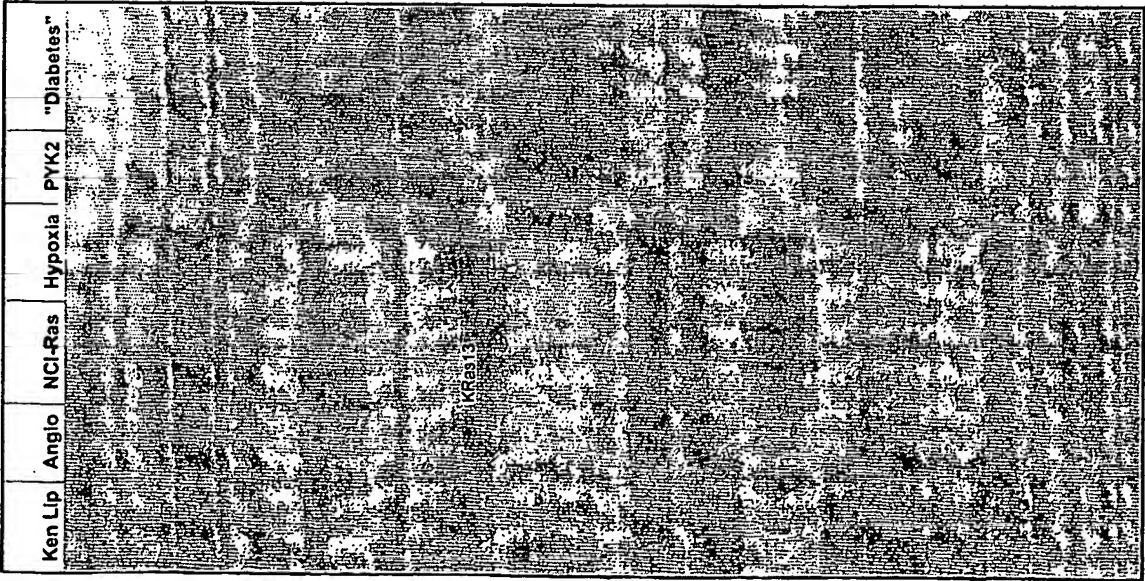


Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
SF539-2							wt	277	145	0	9,827	11	651,624	427
SF539-3							wt	1,599	0	0	21,929	0	547,965	799
SF539-4							wt	0	179	0	19,159	0	449,192	688
SF539-5							wt	85	0	323	23,328	0	516,343	319
SF539-6							wt	0	108	0	7,791	0	682,214	668
SF539-7							wt	0	0	290	17,423	0	574,107	1,461
SF539-8							wt	0	58	0	38,279	0	550,442	1,574
WI 38-1							wt	391	90	25	54,921	0	498,076	672
WI 38-2							wt	725	671	339	93,773	270	339,104	2,179
WI 38-3							wt	1,101	524	65	40,084	0	414,523	945
WI 38-4							wt	6	752	0	37,054	0	620,563	661
WI 38-5							wt	161	289	0	7,149	0	552,397	1,748
WI 38-6							wt	385	63	92	10,543	0	559,003	724
WI 38-7							wt	80	157	0	5,997	0	832,476	1,288
WI 38-8							wt	533	0	0	680	127	1,071,262	1,559
HELA-2h-031899				79				0	0	0	1,827	0	61,571	1,378
HELA-4h-031899				81				0	0	0	174	0	96,166	929
HELA-6h-031899				83				0	232	0	91,313	0	41,904	164
HELA-8h-031899				86				305	0	0	293	0	56,525	291
HELA-10h-031899				90				0	0	0	462	0	80,592	531
HELA-11h-031899				92				0	0	0	2,217	0	50,148	0
HELA-12h-031899				94				496	0	0	231	0	43,735	297
NCI-H322M				146				0	138	0	1,805	0	69,902	1,926
NCI-H460				148				528	0	0	0	432	41,427	0
NOI-H522				150				0	101	0	0	0	64,259	1,627
SNB-19				152				515	0	0	0	217	15,805	214
SNB-75				154				10	0	0	1,207	0	36,729	0
SF-268				156				405	0	0	0	0	11,197	354
SF-295				158				612	630	0	760	532	43,210	1,364
CCRF-CEM				160				0	0	0	599	0	2,830	40
DU-145				162				148	0	0	0	0	55,584	970
HCT 116				164				876	0	0	171	604	48,404	640
Caki-1				166				1,789	63	0	195	357	9,678	0
786-0				168				847	114	0	558	0	5,791	48
T-47D				169				404	55	0	1,859	0	13,446	531
CRL1441 RNA 8/30				181				1,738	0	0	0	0	4,867	0
781T untreated + DNase				183				1,066	16	0	0	590	4,220	168
KB poly A+				194				452	357	0	0	219	43,402	0
HOS poly A+				196				0	0	0	0	904	58,885	468
ACHN				198				426	236	0	193	3,740	67,534	748
UACC-82				200				0	282	0	459	7	12,401	0
MCF-7/ADR-RES				202				776	0	0	209	0	44,167	0
UTOS (Mundy) poly A+				204				0	143	0	373	359	13,394	115
WISH (Collagen) poly A+				206				1,552	0	0	0	0	3,311	0
458 medullo mRNA				208				0	290	0	4,761	0	38,074	0



Tissue type	Release	tag	pos	NC/letc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
SF539-2	R13841	57m	B07	471				SR-10					31					
SF539-3	R13841	57m	B09	472									33					
SF539-4	R13841	57m	B10	473									35					
SF539-5	R13841	57m	B11	474									37					
SF539-6	R13841	57m	B12	475									39					
SF539-7	R13841	57m	B13	476									41					
SF539-8	R13841	57m	B14	477									43					
SF539-9	R13841	57m	B15	478									45					
SF539-10	R13841	57m	B16	479									47					
SF539-11	R13841	57m	B17	480									49					
SF539-12	R13841	57m	B18	481									51					
SF539-13	R13841	57m	B19	482									53					
SF539-14	R13841	57m	B20	483									55					
SF539-15	R13841	57m	B21	484									57					
SF539-16	R13841	57m	B22	485									59					
SF539-17	R13841	57m	B23	486									61					
SF539-18	R13841	57m	B24	487									63					
SF539-19	R13841	57m	B25	488									65					
SF539-20	R13841	57m	B26	489									67					
SF539-21	R13841	57m	B27	490									69					
SF539-22	R13841	57m	B28	491									71					
SF539-23	R13841	57m	B29	492									73					
SF539-24	R13841	57m	B30	493									75					
SF539-25	R13841	57m	B31	494									77					
SF539-26	R13841	57m	B32	495									79					
SF539-27	R13841	57m	B33	496									81					
SF539-28	R13841	57m	B34	497									83					
SF539-29	R13841	57m	B35	498									85					
SF539-30	R13841	57m	B36	499									87					
SF539-31	R13841	57m	B37	500									89					
SF539-32	R13841	57m	B38	501									91					
SF539-33	R13841	57m	B39	502									93					
SF539-34	R13841	57m	B40	503									95					
SF539-35	R13841	57m	B41	504									97					
SF539-36	R13841	57m	B42	505									99					
SF539-37	R13841	57m	B43	506									101					
SF539-38	R13841	57m																

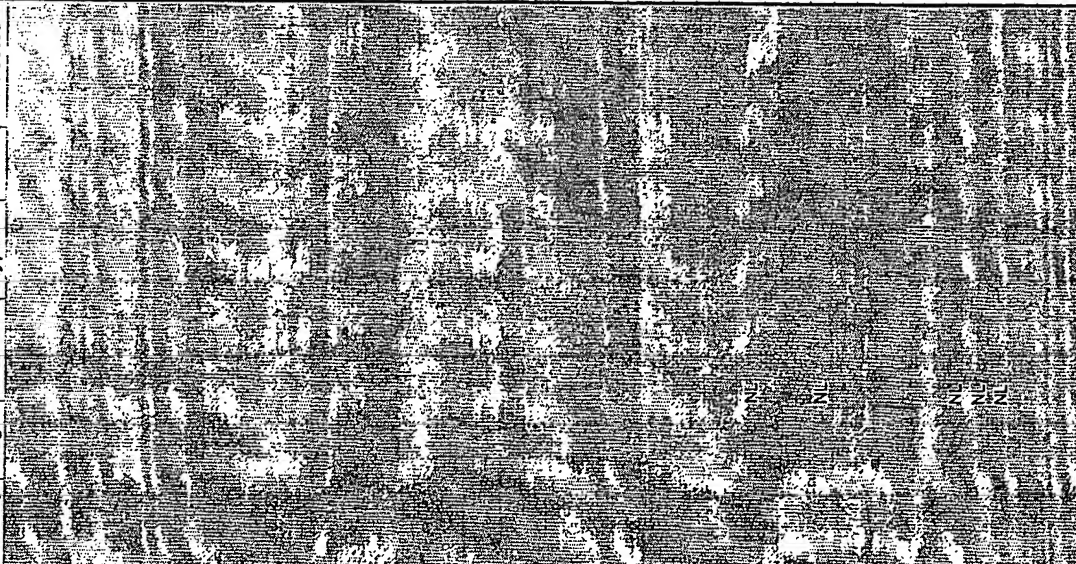
Ken Lip	Anglo	NCI-Ras	Hypoxia	PYK2	"Diabetes"
					

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
CCL137 RNA 3/21/88				218				0	0	0	392	65	35,013	425
WI-38 72h 0.5%FBS. 24h 10% FBS				219				0	0	0	0	0	43,199	0
CRL1441 + TPA (24h) 8/30				220				0	56	0	386	0	13,416	59
HOP-92				241				0	23	0	0	0	22,847	186
MOLT-4				242				0	92	0	309	26	21,250	434
EKVX				243				0	250	0	960	0	77,320	2,875
HL-60				244				0	0	0	908	480	24,392	0
NCL-H23				245				1,129	0	0	211	0	93,263	2,492
RPMI 8226				246				0	0	0	789	0	87,784	1,931
A549/ATCC				247				228	0	0	0	211	5,669	0
SR				248				329	0	0	0	0	15,586	266
OVCAR-3				249				112	0	0	0	0	15,287	240
HCT-15				250				284	387	0	2,254	94	30,611	211
OVCAR-4				251				0	278	0	0	0	4,121	186
UO-31				252				0	0	0	0	0	7,831	96
OVCAR-5				253				181	265	0	1,579	756	28,038	1,658
SN12C				254				0	151	0	0	0	5,978	0
OVCAR-8				255				0	7	0	952	591	67,456	1,954
LOX IMVI				256				239	624	0	2,645	954	25,761	1,176
IGROV1				257				0	0	0	1,111	3	56,485	1,865
SK-MEL-2				258				0	0	0	0	0	2,107	619
SK-OV-3				259				350	0	0	0	0	3,010	95
SK-MEL-5				260				0	0	0	0	0	1,573	0
SF-539				261				726	0	0	889	0	9,243	202
SK-MEL-28				262				148	0	0	1,157	273	23,788	634
K-562				263				313	683	0	763	597	4,477	300
UACC-257				264				245	0	0	26	0	2,255	330
M14				265				90	242	0	0	214	129,614	1,060
MCF7				267				204	0	0	16,366	0	432,526	111
MDA-MB-435				269				90	0	0	1,131	0	14,733	0
HT279				270				0	323	0	0	0	19,273	88
MDA-N				271				12	176	0	89	0	9,135	317
Y79 poly A+				273				463	107	0	0	20	89,877	0
KHOS poly A+				289				0	0	0	0	0	75,507	0
HTB36 24h TPA RNA 6/23				300				5	50	0	0	86	34,541	0
HELA-EXP-031899				313				0	0	0	110	0	55,070	45
HTB36 0h RNA				322				0	20	0	0	0	48,195	54
HT347				323				0	282	0	521	0	46,274	127
458 medulla RNA				324				0	0	0	0	0	26,987	60
NCL-H228				336				214	352	0	0	257	2,862	0
HOP-62				337				219	0	0	0	0	11,386	157
MDA-MB-231				338				0	0	0	0	0	51,134	901
U251				339				0	237	0	577	277	55,589	372
PT cells poly A+				340				130	106	0	4,549	515	9,973	0
PC-3				341				0	284	0	0	261	85,936	1,366
HCC-2998				343				803	0	0	375	639	118,307	958
SW-620				345				0	0	79	22,607	0	467,592	1,300



Table - Tissue Array  
413406\_1.xls

Tissue type	Release	tag	pos	NCI/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
CGI137 RNA3/21/83	R13841	asym	102	77														
WI-38/21/0.5%FBS/24h/10% FBS	R13841	asym	103	86														
CRU1441+TFA (24h) 8.50	R13841	asym	104	78														
HOP-92	R13841	asym	105	1														
MOLT-2	R13841	asym	106	24										242				
EXVA	R13841	asym	107	25										244				
HL-60	R13841	asym	108	26										246				
NCIH23	R13841	asym	109	27										248				
RPMI 8226	R13841	asym	110	34														
A549 ATCC	R13841	asym	111	11														
SR	R13841	asym	112	36														
OVCAR-3	R13841	asym	113	37														
HCT-15	R13841	asym	114	44														
OVCAR-4	R13841	asym	115	46														
UO-3	R13841	asym	116	13														
OVCAR-5	R13841	asym	117	14														
SN12C	R13841	asym	118	15														
OVCAR-8	R13841	asym	119	21														
LOX IMV	R13841	asym	120	23														
IGROV1	R13841	asym	121	25														
SK-MEL-2	R13841	asym	122	26														
SK-MEL-3	R13841	asym	123	27														
SK-MEL-5	R13841	asym	124	28														
SK-MEL-28	R13841	asym	125	29														
K-562	R13841	asym	126	30														
UACC-257	R13841	asym	127	31														
M14	R13841	asym	128	32														
MCF-7-2	R13841	asym	129	33														
MDA-MB-435	R13841	asym	130	34														
MDA-N	R13841	asym	131	35														
MDA-MB-231	R13841	asym	132	36														
MDA-MB-435	R13841	asym	133	37														
MDA-MB-231	R13841	asym	134	38														
MDA-MB-435	R13841	asym	135	39														
MDA-MB-231	R13841	asym	136	40														
MDA-MB-435	R13841	asym	137	41														
MDA-MB-231	R13841	asym	138	42														
MDA-MB-435	R13841	asym	139	43														
MDA-MB-231	R13841	asym	140	44														
MDA-MB-435	R13841	asym	141	45														
MDA-MB-231	R13841	asym	142	46														
MDA-MB-435	R13841	asym	143	47														
MDA-MB-231	R13841	asym	144	48														
MDA-MB-435	R13841	asym	145	49														
MDA-MB-231	R13841	asym	146	50														
MDA-MB-435	R13841	asym	147	51														
MDA-MB-231	R13841	asym	148	52														
MDA-MB-435	R13841	asym	149	53														
MDA-MB-231	R13841	asym	150	54														
MDA-MB-435	R13841	asym	151	55														
MDA-MB-231	R13841	asym	152	56														
MDA-MB-435	R13841	asym	153	57														
MDA-MB-231	R13841	asym	154	58														
MDA-MB-435	R13841	asym	155	59														
MDA-MB-231	R13841	asym	156	60														
MDA-MB-435	R13841	asym	157	61														
MDA-MB-231	R13841	asym	158	62														
MDA-MB-435	R13841	asym	159	63														
MDA-MB-231	R13841	asym	160	64														
MDA-MB-435	R13841	asym	161	65														
MDA-MB-231	R13841	asym	162	66														
MDA-MB-435	R13841	asym	163	67														
MDA-MB-231	R13841	asym	164	68														
MDA-MB-435	R13841	asym	165	69														
MDA-MB-231	R13841	asym	166	70														
MDA-MB-435	R13841	asym	167	71														
MDA-MB-231	R13841	asym	168	72														
MDA-MB-435	R13841	asym	169	73														
MDA-MB-231	R13841	asym	170	74														
MDA-MB-435	R13841	asym	171	75														
MDA-MB-231	R13841	asym	172	76														
MDA-MB-435	R13841	asym	173	77														
MDA-MB-231	R13841	asym	174	78														
MDA-MB-435	R13841	asym	175	79														
MDA-MB-231	R13841	asym	176	80														
MDA-MB-435	R13841	asym	177	81														
MDA-MB-231	R13841	asym	178	82														
MDA-MB-435	R13841	asym	179	83														
MDA-MB-231	R13841	asym	180	84														
MDA-MB-435	R13841	asym	181	85														
MDA-MB-231	R13841	asym	182	86														
MDA-MB-435	R13841	asym	183	87														
MDA-MB-231	R13841	asym	184	88														
MDA-MB-435	R13841	asym	185	89														
MDA-MB-231	R13841	asym	186	90														
MDA-MB-435	R13841	asym	187	91														
MDA-MB-231	R13841	asym	188	92														
MDA-MB-435	R13841	asym	189	93														
MDA-MB-231	R13841	asym	190	94														
MDA-MB-435	R13841	asym	191	95														
MDA-MB-231	R13841	asym	192	96														
MDA-MB-435	R13841	asym	193	97														
MDA-MB-231	R13841	asym	194	98														
MDA-MB-435	R13841	asym	195	99														
MDA-MB-231	R13841	asym	196	100														
MDA-MB-435	R13841	asym	197	101														
MDA-MB-231	R13841	asym	198	102														
MDA-MB-435	R13841	asym	199	103														
MDA-MB-231	R13841	asym	200	104														
MDA-MB-435	R13841	asym	201	105														
MDA-MB-231	R13841	asym	202	106														
MDA-MB-435	R13841	asym	203	107														
MDA-MB-231	R13841	asym	204	108														
MDA-MB-435	R13841	asym	205	109														
MDA-MB-231	R13841	asym	206	110														
MDA-MB-435	R13841	asym	207	111														
MDA-MB-231	R13841	asym	208	112														
MDA-MB-435	R13841	asym	209	113														
MDA-MB-231	R13841	asym	210	114														
MDA-MB-435	R13841	asym	211	115														
MDA-MB-231	R13841	asym	212	116														
MDA-MB-435	R13841	asym	213	117														
MDA-MB-231	R13841	asym	214	118														
MDA-MB-435	R13841	asym	215	119														
MDA-MB-231	R13841	asym	216															

[illegible]

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
HT192				346				0	0	0	0	0	74,630	83
COLO 205				347				621	6	0	0	0	10,126	0
HT218				348				0	0	0	0	0	20,377	0
KM-12				349				860	259	0	823	0	21,590	0
HT151				350				0	411	0	347	0	60,081	0
A498				351				0	428	0	0	378	9,226	414
HT393				352				0	0	0	894	0	57,432	0
RXF 393				353				0	0	0	0	0	6,338	82
TK-10				355				191	0	0	2,314	0	14,179	0
Maine-3M				357				0	0	0	1,531	885	10,560	174
Hs 578T				359				212	0	0	0	9	2,588	0
HT213			50					886	0	0	0	0	48,907	0
HT288			52					0	224	0	683	0	74,176	18
HT139			54					0	0	0	631	0	67,320	102
HT155			56					48	0	0	30	0	61,739	0
HT163			58					0	184	0	773	345	80,453	275
HT170			60					0	0	0	0	0	63,967	0
HT172			62					0	0	0	0	0	48,537	0
HT138			63					10	1	0	741	0	46,416	156
HT178			64					0	14	0	301	0	57,882	330
HT154			65					0	166	0	0	0	69,411	89
HT180			66					47	57	0	0	0	53,278	214
HT169			67					0	0	0	0	0	23,349	0
HT189			68					182	0	0	0	0	63,765	243
HT143			69					0	0	0	86	0	43,069	129
HT190			70					0	0	0	10	0	51,091	22
HT145			71					372	0	0	0	0	58,721	0
HT227			72					101	0	0	173	29	33,093	0
HT302			73					56	207	0	0	0	69,267	275
HT314			74					0	20	0	1,141	0	60,939	287
HT317			76					0	0	0	0	0	51,326	484
Medulloblastoma #425 11/8			77					0	74	0	774	0	76,569	60
HT323			78					0	0	0	0	0	111,765	339
HT327			80					0	251	0	691	432	62,791	338
HT335			82					0	25	0	176	0	69,031	0
HT146			85					227	0	0	0	405	7,160	0
HT348			87					359	0	0	0	0	69,239	1,185
HT311			170					0	75	0	0	0	31,571	246
HT398			185					3,484	281	0	0	259	5,903	31
HT140			187					662	0	0	0	0	1,081	0
HT281			189					549	0	0	0	231	5,431	0
HT1372			191					236	0	0	0	0	11,267	0
TOGP			207					0	0	0	0	265	9,589	16
HT160			216					0	0	0	0	0	2,822	0
HT307			217					144	149	0	0	0	27,812	84
HT369			224					0	103	0	0	0	45,642	0
HT370			226					0	130	0	0	0	33,703	22



Table - Tissue Array  
413406\_1.xls

Tissue type	Release	tag	pos	NCI/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
HT192	R1384	43m	O10	179														
COLO 205	R1384	43m	O11	33														
HT18	R1384	43m	O12	181														
HT112	R1384	43m	O13	135														
HT151	R1384	43m	O14	164														
A498	R1384	43m	O15	38														
HT333	R1384	43m	O16	215														
HT333	R1384	43m	O17	40														
HT110	R1384	43m	O19	43														
HT333	R1384	43m	O21	45														
HT333	R1384	43m	O23	54														
HT213	R1384	43m	O22	180														
HT288	R1384	43m	O24	187														
HT139	R1384	43m	O25	158														
HT155	R1384	43m	O26	166														
HT163	R1384	43m	O27	171														
HT170	R1384	43m	O28	173														
HT172	R1384	43m	O29	174														
HT138	R1384	43m	O30	157														
HT178	R1384	43m	O31	175														
HT154	R1384	43m	O32	165														
HT180	R1384	43m	O33	176														
HT169	R1384	43m	O34	172														
HT189	R1384	43m	O35	177														
HT143	R1384	43m	O36	160														
HT190	R1384	43m	O37	178														
HT145	R1384	43m	O38	161														
HT227	R1384	43m	O39	184														
HT302	R1384	43m	O40	183														
HT314	R1384	43m	O41	193														
HT317	R1384	43m	O42	194														
Meduloblastoma 425	R1384	43m	O43	33														
HT323	R1384	43m	O44	195														
HT327	R1384	43m	O45	196														
HT335	R1384	43m	O46	198														
HT146	R1384	43m	O47	162														
HT348	R1384	43m	O48	201														
HT310	R1384	43m	O49	191														
HT398	R1384	43m	O50	218														
HT140	R1384	43m	O51	159														
HT281	R1384	43m	O52	186														
HT372	R1384	43m	O53	206														
OCGP	R1384	43m	O54	76														
HT160	R1384	43m	O55	169														
HT307	R1384	43m	O56	189														
HT369	R1384	43m	O57	203														
HT370	R1384	43m	O58	204														

Table - Tissue Array  
413406\_1.xls

Ken Lip	Anglo	NCI-Ras	Hypoxia	PYK2	"Diabetes"
[Microarray Data]					



Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor to	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
HT371			228					0	58	0	578	0	48,710	453
HT377			230					0	132	0	0	0	35,666	354
HT382			236					101	328	0	0	0	28,197	45
neuroblastoma RNA			281					1,194	396	0	2,805	0	0	0
HT334			289					0	0	0	0	0	111,643	377
HT338			301					0	56	0	0	0	28,893	0
HT392			315					460	0	0	0	0	19,458	155
HT394			317					0	0	0	46	82	37,089	235
HT312			319					181	36	0	0	245	28,897	0
HT162			325					0	0	0	0	0	17,011	0
HT395			358					233	0	0	0	0	7,637	39
HT157			360					0	383	0	0	0	13,226	0
458 medulla RNA								138	154	0	163	0	80,328	690
CRL1572 3/17/89								0	0	0	248	0	56,547	0
h fibroblasts 3/31/82 #12								423	390	0	0	102	12,530	184
h keratinocytes 2/25/92 #10								64	0	0	0	419	15,538	0
HT308								899	0	0	0	0	54,950	21
HT368								0	31	0	0	80	66,952	0
HT378								0	0	0	40	0	56,777	707
HT385								0	0	0	846	0	33,000	0

Table - Tissue Array  
413406\_1.xls

Tissue_type	Release	tag	pos	NC/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
HT371	R1384-I	asym	J12	205														
HT377	R1384-I	asym	J14	208														
HT382	R1384-I	asym	J20	210														
nanoblastoma RNA	R1384-I	asym	L17	84														
HT334	R1384-I	asym	M1	197														
HT338	R1384-I	asym	M13	199														
HT392	R1384-I	asym	N03	213														
HT394	R1384-I	asym	N05	216														
HT312	R1384-I	asym	N07	192														
HT182	R1384-I	asym	N13	170														
HT395	R1384-I	asym	O22	217														
HT157	R1384-I	asym	O24	167														
438 medulo RNA	R1384-I	asym	B21	282														
CR1572 3/17/89	R1384-I	asym	D03	80														
fibroblast 3/3/92 #12	R1384-I	asym	L21	132														
keratinocytes 2/25/92 #10	R1384-I	asym	J16	133														
HT308	R1384-I	asym	D23	190														
HT388	R1384-I	asym	D17	202														
HT378	R1384-I	asym	D19	209														
HT385	R1384-I	asym	D21	212														

Table - Tissue Array  
413406\_1.xls

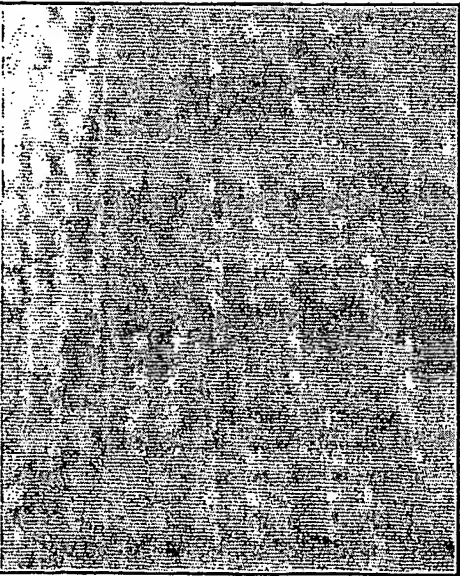
Ken Lip	Anglo	NCI-Ras	Hypoxia	PYK2	"Diabetes"
					

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
TCGP	26							212	0	0	2,366	0	58,451	2,077
h keratinocytes 2/25/92 #10	46							83	3,050	0	2,340	0	89,541	1,233
h adult SMC 10/21/92 #17	47							0	0	0	1,713	374	73,933	291
h fibroblasts 3/31/92 #12	48							0	0	0	423	152	62,710	1,235
HOP-92	97							514	0	0	2,631	94	87,584	1,316
OVCAR-8	98							16	0	0	0	0	4,204	164
ERXV	99							19	0	0	231	0	27,721	105
IGROV1	100							0	0	0	65	0	7,800	0
NCI-H23	101							0	0	0	0	0	26,620	70
SK-OV-3	102							260	0	0	1,474	0	65,008	1,516
NCI-H226	103							153	365	0	1,091	91	107,932	3,834
SNB-19	104							0	0	0	855	0	74,412	1,302
NCI-H322M	105							243	77	0	422	226	35,551	84
SNB-75	106							138	0	0	0	0	98,053	1,797
NCI-H460	107							0	548	0	3,348	81	78,829	1,166
U251	108							0	554	0	0	0	84,310	677
NCI-H522	109							196	0	0	182	0	22,554	497
SF-268	110							0	771	0	0	593	49,287	1,115
AS49/ATCC	111							0	268	0	1,613	1,619	43,786	1,448
SF-295	112							0	0	0	1,333	0	7,106	0
HOP-62	113							231	343	0	529	1,746	21,968	670
SF-539	114							176	101	0	237	167	39,588	1,662
OVCAR-3	115							0	179	0	998	442	28,838	737
CCRF-CEM	116							80	6,816	0	605	0	45,975	2,130
OVCAR-4	117							52	57	0	704	292	34,439	1,510
K-562	118							0	0	0	0	185	33,123	1,072
OVCAR-5	119							0	0	0	1,307	542	18,164	23
MOLT-4	120							0	0	0	118	133	35,133	410
HL-60	121							0	537	0	0	659	80,247	719
SN12C	122							32	0	0	0	711	76,151	1,579
RPMI 8226	123							0	0	0	0	0	15,177	419
A498	124							0	904	0	430	3,358	122,082	962
SR	125							80	0	0	0	553	97,140	1,864
Calu-1	126							560	0	0	1,099	6,386	88,518	1,978
DU-145	127							51	191	0	169	0	3,687	0
RXF 393	128							0	202	0	685	1,192	108,097	1,825
PC-3	129							0	0	0	0	55	27,208	99
ACHN	130							0	0	0	1,248	0	21,998	0
HCC-2998	131							0	0	0	0	0	14,182	0
786-0	132							489	443	0	500	302	74,251	689
HCT 116	133							94	149	0	397	0	2,376	0
Tk-10	134							503	102	0	1,682	415	51,674	1,842
SW-620	135							865	25	148	21,381	90	476,097	853
LOX IMVI	136							608	348	0	0	610	8,504	241
COLO 205	137							134	260	0	223	628	60,452	610
Melme-3M	138							69	281	0	416	734	33,878	1,065
HCT-15	139							0	0	0	0	781	35,146	562

[illegible]



Table - Tissue Array  
413406\_1.xls

Ken Lip	Anglo	MCI-Ras	Hypoxia	PK2	"Diabetes"
		NL			
		NL			
		NL			
		NL			
		Kras			
		NL			
		NL			
		NL			
		NL			
		NL			
		Kras12			
		Kras12			
		Kras12			
		NL			
		NL			
		NL			
		Kras 2			
		NL			
		Kras13			

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
SK-MEL-2	140							720	219	0	1,523	340	21,933	646
SK-MEL-3	141							0	227	0	0	0	23,168	0
SK-MEL-5	142							346	1,100	0	1,862	319	20,650	825
UO-31	143							1,235	0	0	0	0	43,858	549
SK-MEL-28	144							0	121	0	362	151	2,283	158
UACC-62	145							0	74	0	473	1,045	106,158	1,533
UACC-257	147							1,027	131	0	1,889	0	132,416	2,031
M14	149							0	182	0	67	0	14,280	59
MCF7	151							66	0	106	17,887	0	476,093	791
MCF-7/ADR-RES	153							0	0	0	972	484	87,261	1,487
Hs 578T	155							0	0	0	1,717	0	125,561	611
MDA-MB-231	157							0	416	0	0	107	77,227	0
MDA-MB-435	159							391	359	0	1,578	0	41,654	1,257
MDA-N	161							0	183	0	2,141	360	28,371	1,457
T-47D	163							139	103	0	0	202	52,411	220
adrenal gland - h		1						64	29	0	16,403	133	808,208	1,824
lymph node - h		2						0	856	0	2,522	0	126,297	6,180
bone marrow - h		3						0	793	0	157	435	127,325	3,459
mammary gland - h		4						0	0	0	0	221	60,842	482
brain - h		5						0	309	0	206	884	109,916	4,168
pancreas - h		6						0	400	0	1,126	0	103,318	2,814
cerebellum - h		7						0	87	0	9,410	578	187,028	9,694
pituitary gland - h		8						83	1,273	0	1,820	194	108,893	5,097
fetal brain - h		9						359	1,004	0	4,828	862	121,604	8,651
placenta - h		10						0	0	0	6,212	1,310	115,612	2,852
fetal kidney - h		11						0	0	0	4,629	28,058	192,142	4,075
prostate, h		12						0	290	0	851	0	18,517	0
fetal liver - h		13						5,768	69	0	2,652	7,321	128,455	2,039
salivary gl. - h		14						0	0	0	1,389	0	92,749	1,033
fetal lung - h		15						0	666	0	3,637	1,462	198,037	4,592
skeletal muscle - h		16						1,336	0	0	2,995	0	43,019	42
heart - h		17						12,477	125	0	2,160	205	107,943	899
small intestine - h		18						0	228	0	1,736	1,221	117,752	882
kidney - h		19						0	56	0	2,418	18,839	135,359	1,267
spinal cord - h		20						0	144	0	1,031	451	119,848	789
liver - h		21						379	1,481	0	433	0	153,048	852
Spleen - h		22						778	191	0	861	0	122,519	1,248
lung - h		23						0	530	0	1,784	0	76,316	733
stomach - h		24						372	0	0	171	0	39,162	278
testis - h		25						15	0	0	14,003	0	141,552	5,293
thymus - h		27						1,943	343	0	2,504	0	180,985	6,488
HPAEC		28				28		0	0	0	2,192	0	171,007	2,973
thyroid gland - h		29						0	561	0	578	0	155,558	2,783
RPTC		30				30		166	170	0	2,416	0	40,468	0
trachea - h		31						0	5	0	2,843	618	126,079	3,124
HMEC		32						0	0	0	0	0	14,267	0
uterus - h		33						0	492	0	2,700	171	120,188	3,561

Tissue_type	Release	tag	pos	NCI/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
SK-MEL-2	R1384	sym	F20	269														
KM-12	R1384	sym	F21	258														
SK-MEL-5	R1384	sym	F22	270														
UO-3	R1384	sym	F23	259														
SK-MEL-28	R1384	sym	F24	271														
UACC-62	R1384	sym	G01	272														
UACC-257	R1384	sym	G03	273														
MCF-7	R1384	sym	G05	274														
MCF-7	R1384	sym	G07	275														
MCF-7	R1384	sym	G09	276														
Hs578T	R1384	sym	G11	277														
MDAMB-231	R1384	sym	G13	278														
MDAMB-435	R1384	sym	G15	279														
MDAMB-435	R1384	sym	G17	280														
MDAMB-435	R1384	sym	G19	281														
ADRRES-2	R1384	sym	A01	285														
ADRRES-2	R1384	sym	A02	297														
bone marrow h	R1384	sym	A03	286														
mammary gland h	R1384	sym	A04	298														
brain h	R1384	sym	A05	287														
pancreas h	R1384	sym	A06	299														
cerecudum h	R1384	sym	A07	288														
pituitary gland h	R1384	sym	A08	100														
testis h	R1384	sym	A09	289														
fetal brain h	R1384	sym	A10	301														
placenta h	R1384	sym	A11	300														
fetal kidney h	R1384	sym	A12	302														
prostate h	R1384	sym	A13	291														
fetal liver h	R1384	sym	A14	303														
Salivary gl. h	R1384	sym	A15	292														
retallung h	R1384	sym	A16	304														
Skeletal muscle h	R1384	sym	A17	293														
heart h	R1384	sym	A18	305														
small intestine h	R1384	sym	A19	294														
Mooney's D	R1384	sym	A20	306														
spinal cord h	R1384	sym	A21	295														
Brain h	R1384	sym	A22	307														
Spleen h	R1384	sym	A23	296														
lung h	R1384	sym	A24	309														
stomach h	R1384	sym	B01	309														



**Table - Tissue Array**  
**413406\_1.xls**

[illegible]

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor to	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
HCAEC		34						495	0	0	0	300	11,188	0
Pancreas - h		35						0	87	0	482	0	29,237	0
lymph node - h		36						326	298	0	0	0	16,454	0
Skeletal muscle - h		37						242	0	0	1,216	0	164,981	412
fetal liver - h		38						0	0	0	506	0	135,364	2,681
Heart - h		39						500	94	0	2,052	0	84,824	394
thymus.h		40						717	166	0	145	44	68,060	2,415
Duodenum - h		41						490	113	0	2,188	421	70,179	559
Fetal brain - h		42						189	0	0	0	0	32,869	359
Salivary gl. - h		43						0	221	0	396	0	45,754	196
testis - h		44						0	0	0	0	0	83,131	273
fetal kidney - h								1,124	0	0	0	0	67,258	423
fetal liver - h		51		49				443	0	0	0	0	18,565	208
fetal lung - h		53		53				0	109	0	1,903	89	32,265	251
heart - h		55		55				0	45	0	288	0	23,292	280
kidney - h		57		57				0	0	0	45	0	84,812	1,129
lung - h		59		59				0	183	0	0	0	80,505	199
lymph node - h		61		61				45	0	0	0	0	143,506	3,584
WI-38 72h		179		179				0	149	0	0	0	17,291	327
thymus - h		193		193				0	0	0	0	659	60,103	252
HEPM 3d TGFβ1 detergent+DNase		195		195				69	134	0	0	0	26,232	166
Salivary gl. - h		197		197				0	0	0	278	0	11,196	143
testis - h		199		199				0	86	0	98	0	32,630	0
Pancreas - h		201		201				199	0	0	0	0	50,700	0
Skeletal muscle - h		203		203				0	0	0	0	32	64,559	197
Duodenum - h		205		205				0	0	0	0	0	33,654	0
HMEC		209		209				777	102	0	0	206	101,094	3,485
fetal brain - h		210		210				0	69	0	1,050	0	26,545	248
HCAEC		211		211				0	210	0	289	313	106,354	2,700
placenta - h		212		212		211		0	296	0	0	0	16,093	223
fetal liver - h		213		213				1,358	40	0	0	345	47,791	176
stomach - h		214		214				0	0	0	151	0	26,452	196
Heart - h		215		215				0	0	0	92	0	51,494	0
Bladder - h		222		222				0	140	0	0	0	39,930	0
HT372-normal		234		234				0	50	0	0	0	19,299	10
HT382-normal		266		266				615	78	0	0	0	86,365	202
HT392-normal		268		268				131	0	0	0	0	30,692	361
HPAEC		275		275		275		2,553	446	166	583	329	1,532,778	4,110
adrenal gland - h		277		277				0	0	0	0	0	40,511	422
bone marrow - h		279		279				0	19	0	0	0	6,744	114
skeletal muscle - h		290		290				0	45	0	1,545	0	102,189	1,579
small intestine - h		292		292				0	0	0	0	0	40,631	318
spinal cord - h		294		294				0	0	0	357	0	48,103	502
Spleen - h		296		296				0	3	0	657	153	71,535	60
liver - h		297		297				375	0	0	0	0	88,510	74
testis - h		298		298				422	0	0	640	0	167,195	2,423
bladder - h		302		302				38	311	0	2,206	0	74,917	825

Table - Tissue Array  
413406\_1.xls

Tissue_type	Release	tag	pos	NCI/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
HCAEC	R13841	sym	B10	327														
pancreas	R13841	sym	B11	314														
lymph node	R13841	sym	B12	325														
skeletal muscle	R13841	sym	B13	315														
colon	R13841	sym	B14	324														
liver	R13841	sym	B15	316														
heart	R13841	sym	B16	326														
thymus	R13841	sym	B17	317														
duodenum	R13841	sym	B18	326														
fetal brain	R13841	sym	B19	318														
salivary gland	R13841	sym	B20	327														
testis	R13841	sym	C01	97														
fetal kidney	R13841	sym	C03	93														
fetal liver	R13841	sym	C05	94														
heart	R13841	sym	C07	95														
kidney	R13841	sym	C09	96														
lung	R13841	sym	G11	98														
lymph node	R13841	sym	G13	99														
WI-38/2h	R13841	sym	H11	85														
thymus	R13841	sym	I01	112														
HEP-2d	R13841	sym	I03	70														
salivary gland	R13841	sym	I05	126														
testis	R13841	sym	I07	125														
pancreas	R13841	sym	I09	116														
Skeletal muscle	R13841	sym	I11	117														
duodenum	R13841	sym	I13	119														
HMEC	R13841	sym	I17	123														
Fetal brain	R13841	sym	I18	91														
HCAEC	R13841	sym	I20	124														
placenta	R13841	sym	I21	103														
fetal liver	R13841	sym	I22	110														
stomach	R13841	sym	I23	118														
heart	R13841	sym	I26	130														
bladder	R13841	sym	I27	120														
HT132 normal	R13841	sym	I28	106														
HT132 normal	R13841	sym	I29	107														
HT132 normal	R13841	sym	I30	108														
HPAEC	R13841	sym	I31	109														
MDR-RES	R13841	sym	I32	110														
bone marrow	R13841	sym	I33	111														
skeletal muscle	R13841	sym	I34	112														
small intestine	R13841	sym	I35	113														
spinal cord	R13841	sym	I36	114														
spleen	R13841	sym	I37	115														
liver	R13841	sym	I38	116														
pancreas	R13841	sym	I39	117														
bladder	R13841	sym	I40	118														

Table - Tissue Array  
413406\_1.xls

Ken Lip	Angio	NCI-Ras	Hypoxia	PKK2	"Diabetes"
34	34				35
38					37
49	49				
51	51				
53	53				
					20
					203
211	211				
212	212				
213	213				
275	275				
					280

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
mammary gland - h					303			221		0	794	117	48,744	0
pancreas - h					305			678	16	0	817	473	127,476	342
pituitary gland - h					307			0	0	0	0	0	20,657	0
prostate, h					309			0	634	0	297	2,243	148,345	3,258
salivary gl. - h					311			455	0	0	1,109	0	18,958	0
thyroid gland - h					314			0	0	0	0	0	127,246	72
trachea - h					316			0	125	0	16	172	80,358	0
uterus - h					318			19	0	0	1,641	0	73,044	1,084
HEPM 3d untreated					320			1,065	18	0	0	510	61,981	0
HT149 - normal					321			528	0	0	0	0	56,128	155
thymus, h					326			0	0	0	0	0	51,419	0
HT398-normal					327			551	431	0	742	0	32,502	24
Fetal brain - h					328			734	0	0	1,133	0	166,205	3,822
h adult SMC 10/21/92 #17					330			0	0	0	0	0	23,012	164
lymph node - h					332			95	0	0	0	67	22,358	0
RPTC					334	334		1,186	189	0	223	0	96,951	2,705
brain - h					342			0	204	0	0	716	51,925	0
cerebellum - h					344			0	0	0	40	0	31,053	0
HT157-normal					361			0	285	0	0	0	81,782	1
HT213-normal					363			1,947	45	0	15	0	20,122	0
HT218-normal					365			0	0	0	0	0	13,192	0
HeLa - 1								188	27	0	17,506	191	555,591	1,745
HeLa - 2								362	758	2	52,193	0	679,609	1,613
HeLa - 3								1,051	0	67	19,138	0	443,018	1,179
HeLa - 4								699	387	0	11,837	0	636,546	1,133
HeLa - 5								1,571	0	600	40,632	0	740,204	1,934
HeLa - 6								0	215	0	40,279	174	849,472	3,502
HeLa - 7								233	0	183	39,006	81	914,876	2,315
HeLa - 8								0	0	0	30,268	0	967,084	1,982
ADR-RES - 1								540	168	0	48,622	0	921,921	922
ADR-RES - 2								76	0	0	18,959	0	1,010,246	591
ADR-RES - 3								334	344	0	13,658	0	359,976	383
ADR-RES - 4								0	0	0	8,302	0	442,971	11
ADR-RES - 5								306	0	164	20,910	0	544,895	130
ADR-RES - 6								140	0	0	26,446	0	601,848	468
ADR-RES - 7								0	661	0	1,745	429	85,849	3,468
ADR-RES - 8								0	283	0	3,363	0	456,295	578
C33A - 1								0	0	0	3,630	0	389,598	609
C33A - 2								0	0	0	13,815	0	379,820	824
C33A - 3								4,866	0	104	9,604	0	498,796	1,112
C33A - 4								89	0	128	8,326	42	644,383	890
C33A - 5								0	0	0	34,820	0	678,800	0
C33A - 6								0	0	0	690	0	785,878	908
C33A - 7								0	0	149	18,136	0	1,172,556	2,189
C33A - 8								21	0	173	42,512	0	269,251	221
EKVX - 1								675	0	0	27,186	0	498,108	0
EKVX - 2								0	0	92	69,264	272	498,108	0



Table - Tissue Array  
413406\_1.xls

Tissue_type	Release	tag	pos	NCI/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
mammary gland: h	R1384-1	35ym	M15	100														
pancreas: h	R1384-1	35ym	M17	101														
pituitary gland: h	R1384-1	35ym	M19	102														
prostate: h	R1384-1	35ym	M21	104														
SA-OS (Mundy) pol/A+	R1384-1	35ym	M23	105								308						
thyroid gland: h	R1384-1	35ym	N02	113														
trachea: h	R1384-1	35ym	N04	114														
testis: h	R1384-1	35ym	N06	115														
MEPM 3d untreated	R1384-1	35ym	N08	171														
HT179: normal	R1384-1	35ym	N09	163														
lymph node: h	R1384-1	35ym	N14	177														
HT358: normal	R1384-1	35ym	N15	219														
Fetal brain: h	R1384-1	35ym	N16	126														
adult SMC 10/27/92 #17	R1384-1	35ym	N18	175									328					
lymph node: h	R1384-1	35ym	N20	175														
R1384-1	R1384-1	35ym	N22	122														
brain: h	R1384-1	35ym	O06	89														
carcinoma: h	R1384-1	35ym	O08	90									342					
HT157: normal	R1384-1	35ym	P01	169									344					
HT213: normal	R1384-1	35ym	P03	181														
HT218: normal	R1384-1	35ym	P05	183														
HeLa-1	R2134-1	35ym	G10	184	CC-26		MC-31	SR-21						58		58	58	58
HeLa-2	R2134-1	35ym	F18	395	CC-27			SR-22						138				
HeLa-3	R2134-1	35ym	G12	396	CC-27											60	60	60
HeLa-4	R2134-1	35ym	G14	397	CC-28											62	62	62
HeLa-5	R2134-1	35ym	C16	398	CC-29											64	64	64
HeLa-6	R2134-1	35ym	C18	399	CC-30													
HeLa-7	R2134-1	35ym	G20	400	CC-30		MC-32									68	68	68
HeLa-8	R2134-1	35ym	G22	401	CC-30		MC-33									70	70	70
ADR-RES-3	R2134-1	35ym	C17	341	CC-16		MC-28	SR-19						65		65	65	65
ADR-RES-4	R2134-1	35ym	C19	342	CC-17		MC-29	SR-20						67		67	67	67
ADR-RES-5	R2134-1	35ym	G21	343	CC-17													
ADR-RES-6	R2134-1	35ym	G23	344	CC-18											69	69	69
ADR-RES-7	R2134-1	35ym	G24	345	CC-19											71	71	71
ADR-RES-8	R2134-1	35ym	G26	346	CC-20											50	50	50
adrenal gland: h	R2134-1	35ym	O06	347	CC-20		MC-29									54	54	54
adrenal gland: h	R2134-1	35ym	O08	348	CC-20		MC-30									56	56	56
C33A-1	R2134-1	35ym	D08	349	CC-20		MC-40	SR-27						80		80	80	80
C33A-2	R2134-1	35ym	D10	350	CC-20			SR-28						82		82	82	82
C33A-3	R2134-1	35ym	D12	351	CC-20													
C33A-4	R2134-1	35ym	D14	352	CC-20													
C33A-5	R2134-1	35ym	D16	353	CC-20													
C33A-6	R2134-1	35ym	D18	354	CC-20													
C33A-7	R2134-1	35ym	D20	355	CC-20		MC-41									92	92	92
C33A-8	R2134-1	35ym	D22	356	CC-20		MC-42									84	84	84
EVX-1	R2134-1	35ym	F03	370	CC-06		MC-04	SR-03						122		122	122	122
EVX-2	R2134-1	35ym	F04	371	CC-06			SR-04						124		124	124	124

Table - Tissue Array  
413406\_1.xls

Ken Lip	Angio	NCI-Ras	Hypoxia	PYK2	"Diabetes"
305					
304					

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor to	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK386 ID#NA_43
EKVX-3							mutant	0	4	0	93,678	0	429,070	340
EKVX-4							mutant	577	0	71	140,696	344	429,429	819
EKVX-5							mutant	29	0	0	34,966	0	813,745	1,003
EKVX-6							mutant	6	0	98	7,689	79	459,086	799
EKVX-7							mutant	705	0	564	41,327	383	1,353,897	2,803
EKVX-8							mutant	192	242	0	31,342	0	397,494	1,472
H1299-1							mutant	1,259	0	62	26,353	0	501,247	1,055
H1299-2							mutant	0	327	37	13,882	0	712,338	1,004
H1299-3							mutant	0	121	16	36,682	0	778,091	2,440
H1299-4							mutant	172	0	22	57,330	669	957,276	2,317
H1299-5							mutant	0	0	92	18,628	0	865,896	1,039
H1299-6							mutant	0	79	52	17,264	172	338,356	412
H1299-7							mutant	147	0	0	12,801	0	523,628	0
H1299-8							mutant	2,032	0	0	21,626	553	519,154	449
H129-1							mutant	817	0	85	60,106	0	859,216	1,413
H129-2							mutant	1,145	298	16	62,721	0	440,697	1,568
H129-3							mutant	833	228	132	85,651	228	700,898	958
H129-4							mutant	120	0	258	102,716	604	1,046,464	1,512
H129-5							mutant	810	0	0	137,429	0	1,522,264	3,276
H129-6							mutant	3,401	128	154	131,385	568	1,157,087	2,892
H129-7							mutant	0	0	25	21,556	455	359,692	523
H129-8							mutant	0	0	227	34,495	11	539,304	1,099
OVCAR-5-1							mutant	166	0	0	46,515	403	407,563	1,031
OVCAR-5-2							mutant	0	0	81	22,899	15	550,448	653
OVCAR-5-3							mutant	66	344	186	67,219	0	386,258	33
OVCAR-5-4							mutant	0	81	0	28,479	1	480,438	2,500
OVCAR-5-5							mutant	16	314	0	52,491	140	607,764	883
OVCAR-5-6							mutant	0	144	41	5,128	0	652,947	91
OVCAR-5-7							mutant	1,388	743	379	73,397	0	2,126,515	1,252
OVCAR-5-8							mutant	687	0	0	72,830	1,009	411,234	1,283
SF-268-1							mutant	454	0	0	48,444	169	1,247,938	3,189
SF-268-2							mutant	175	0	1	31,761	0	294,676	519
SF-268-3							mutant	35	0	211	26,186	0	442,870	212
SF-268-4							mutant	178	184	204	51,244	0	399,799	568
SF-268-5							mutant	0	816	0	56,103	0	471,620	0
SF-268-6							mutant	0	36	0	25,570	0	394,514	619
SF-268-7							mutant	850	143	0	40,201	429	347,362	1,837
SF-268-8							mutant	2,391	138	160	67,712	12	483,334	1,254
SW480-1							mutant	0	0	0	26	0	13,937	30
SW480-2							mutant	0	255	0	1,000	79	33,878	1,032
SW480-3							mutant	470	0	0	76,090	0	1,460,381	3,005
SW480-4							mutant	79	0	0	45,130	226	423,651	537
SW480-5							mutant	260	85	206	33,127	0	392,824	235
SW480-6							mutant	733	0	0	11,557	0	536,855	116
SW480-7							mutant	1,608	0	0	38,957	0	729,185	983
SW480-8							mutant	183	0	0	27,056	0	555,267	1,473
U2OS-1							mutant	1,263	125	0	78,276	1,905	1,630,846	4,317



Table - Tissue Array  
413406\_1.xls

Tissue type	Release	tag	pos	NCI/cct	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
EVX-3	R21841	sym	F06	372	CC-08	DD-05											128	128
EVX-4	R21841	sym	F08	373	CC-07	DD-05											128	128
EVX-5	R21841	sym	A17	374	CC-08												17	19
EVX-6	R21841	sym	A19	375		DD-08											21	
EVX-7	R21841	sym	A2	376	CC-10		MC-05										21	
EVX-8	R21841	sym	F10	377			MC-06										130	
EVX-9	R21841	sym	D15	378	CC-31		MC-07	SR-25									87	87
EVX-10	R21841	sym	D17	379				SR-26									87	
EVX-11	R21841	sym	D19	380	CC-32												87	
EVX-12	R21841	sym	D21	381	CC-33												87	
EVX-13	R21841	sym	D23	382	CC-34												87	
EVX-14	R21841	sym	D25	383													87	
EVX-15	R21841	sym	D27	384	CC-35		MC-08										87	
EVX-16	R21841	sym	D29	385			MC-09										87	
EVX-17	R21841	sym	D31	386			MC-10										87	
EVX-18	R21841	sym	D33	387													87	
EVX-19	R21841	sym	D35	388													87	
EVX-20	R21841	sym	D37	389													87	
EVX-21	R21841	sym	D39	390													87	
EVX-22	R21841	sym	D41	391													87	
EVX-23	R21841	sym	D43	392													87	
EVX-24	R21841	sym	D45	393													87	
EVX-25	R21841	sym	D47	394													87	
EVX-26	R21841	sym	D49	395													87	
EVX-27	R21841	sym	D51	396													87	
EVX-28	R21841	sym	D53	397													87	
EVX-29	R21841	sym	D55	398													87	
EVX-30	R21841	sym	D57	399													87	
EVX-31	R21841	sym	D59	400													87	
EVX-32	R21841	sym	D61	401													87	
EVX-33	R21841	sym	D63	402													87	
EVX-34	R21841	sym	D65	403													87	
EVX-35	R21841	sym	D67	404													87	
EVX-36	R21841	sym	D69	405													87	
EVX-37	R21841	sym	D71	406													87	
EVX-38	R21841	sym	D73	407													87	
EVX-39	R21841	sym	D75	408													87	
EVX-40	R21841	sym	D77	409													87	
EVX-41	R21841	sym	D79	410													87	
EVX-42	R21841	sym	D81	411													87	
EVX-43	R21841	sym	D83	412													87	
EVX-44	R21841	sym	D85	413													87	
EVX-45	R21841	sym	D87	414													87	
EVX-46	R21841	sym	D89	415													87	
EVX-47	R21841	sym	D91	416													87	
EVX-48	R21841	sym	D93	417													87	
EVX-49	R21841	sym	D95	418													87	
EVX-50	R21841	sym	D97	419													87	
EVX-51	R21841	sym	D99	420													87	
EVX-52	R21841	sym	D101	421													87	
EVX-53	R21841	sym	D103	422													87	
EVX-54	R21841	sym	D105	423													87	
EVX-55	R21841	sym	D107	424													87	
EVX-56	R21841	sym	D109	425													87	
EVX-57	R21841	sym	D111	426													87	
EVX-58	R21841	sym	D113	427													87	
EVX-59	R21841	sym	D115	428													87	
EVX-60	R21841	sym	D117	429													87	
EVX-61	R21841	sym	D119	430													87	
EVX-62	R21841	sym	D121	431													87	
EVX-63	R21841	sym	D123	432													87	
EVX-64	R21841	sym	D125	433													87	
EVX-65	R21841	sym	D127	434													87	
EVX-66	R21841	sym	D129	435													87	
EVX-67	R21841	sym	D131	436													87	
EVX-68	R21841	sym	D133	437													87	
EVX-69	R21841	sym	D135	438													87	
EVX-70	R21841	sym	D137	439													87	
EVX-71	R21841	sym	D139	440													87	
EVX-72	R21841	sym	D141	441													87	
EVX-73	R21841	sym	D143	442													87	
EVX-74	R21841	sym	D145	443													87	
EVX-75	R21841	sym	D147	444													87	
EVX-76	R21841	sym	D149	445													87	
EVX-77	R21841	sym	D151	446													87	
EVX-78	R21841	sym	D153	447													87	
EVX-79	R21841	sym	D155	448													87	
EVX-80	R21841	sym	D157	449													87	
EVX-81	R21841	sym	D159	450													87	
EVX-82	R21841	sym	D161	451													87	
EVX-83	R21841	sym	D163	452													87	
EVX-84	R21841	sym	D165	453													87	
EVX-85	R21841	sym	D167	454													87	
EVX-86	R21841	sym	D169	455													87	
EVX-87	R21841	sym	D171	456													87	
EVX-88	R21841	sym	D173	457													87	
EVX-89	R21841	sym	D175	458													87	
EVX-90	R21841	sym	D177	459													87	
EVX-91	R21841	sym	D179	460													87	
EVX-92	R21841	sym	D181	461													87	
EVX-93	R21841	sym	D183	462													87	
EVX-94	R21841	sym	D185	463													87	
EVX-95	R21841	sym	D187	464													87	
EVX-96	R21841	sym	D189	465													87	
EVX-97	R21841	sym	D191	466													87	
EVX-98	R21841	sym	D193	467													87	
EVX-99	R21841	sym	D195	468													87	
EVX-100	R21841	sym	D197	469													87	
EVX-101	R21841	sym	D199	470													87	
EVX-102	R21841	sym	D201	471													87	
EVX-103	R21841	sym	D203	472													87	
EVX-104	R21841	sym	D205	473													87	
EVX-105	R21841	sym	D207	474													87	
EVX-106	R21841	sym	D209	475													87	
EVX-107	R21841	sym	D211	476													87	
EVX-108	R21841	sym	D213	477													87	
EVX-109	R21841	sym	D215	478													87	
EVX-110	R21841	sym	D217	479													87	
EVX-111	R21841	sym	D219	480													87	
EVX-112	R21841	sym	D221	481													87	
EVX-113	R21841	sym	D223	482													87	
EVX-114	R21841	sym	D225	483													87	
EVX-115	R21841	sym	D227	484													87	
EVX-116	R21841	sym	D229	485														

Table - Tissue Array  
413406\_1.xls

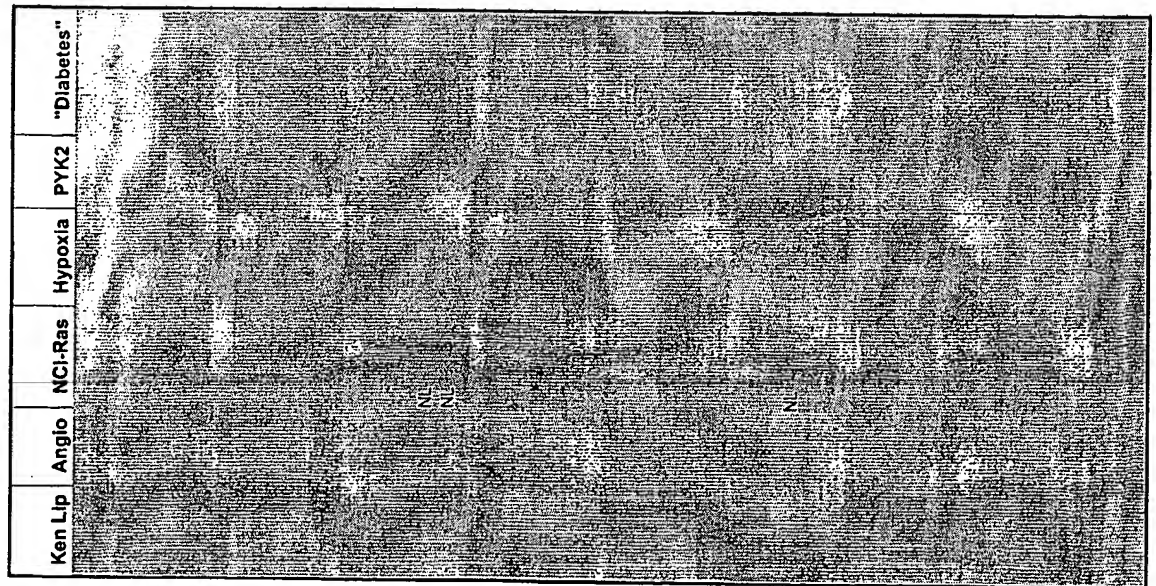
Ken Lip	Angio	NCI-Ras	Hypoxia	PYK2	"Diabetes"
					

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
U2OS - 2							mutant	641	327	202	29,915	0	432,337	872
U2OS - 3							mutant	573	0	1	40,362	127	406,461	412
U2OS - 4							mutant	1,545	0	182	10,515	106	398,039	738
U2OS - 5							mutant	679	0	94	37,160	0	491,594	827
U2OS - 6							mutant	0	294	108	5,845	3,862	371,040	1,175
U2OS - 7							mutant	0	0	0	22,921	17	500,353	1,551
U2OS - 8							mutant	0	0	0	16,035	0	472,901	1,978
A549 - 1							wt	0	0	112	54,563	0	546,984	523
A549 - 2							wt	1,553	0	0	25,728	192	490,791	367
A549 - 3							wt	0	0	0	8,583	0	526,257	173
A549 - 4							wt	58	118	92	27,986	170	637,978	439
A549 - 5							wt	0	119	0	56,834	0	470,171	657
A549 - 6							wt	0	0	203	2,773	0	560,302	671
A549 - 7							wt	240	0	188	32,976	0	656,478	1,455
A549 - 8							wt	57	157	1	74,102	243	409,754	1,458
HCT-116 - 1							wt	0	0	0	100,001	85	2,916,547	3,159
HCT-116 - 2							wt	0	16	4	145,047	0	386,158	1,870
HCT-116 - 3							wt	665	0	0	17,785	159	427,289	313
HCT-116 - 4							wt	275	0	0	38,703	0	386,481	198
HCT-116 - 5							wt	556	179	38	37,110	85	505,899	955
HCT-116 - 6							wt	754	0	105	61,500	0	498,369	488
HCT-116 - 7							wt	190	187	47	42,722	0	416,088	1,089
HCT-116 - 8							wt	783	0	49	48,907	133	555,183	1,362
Hs68 - 1							wt	0	0	50	21,139	0	412,159	1,066
Hs68 - 2							wt	0	84	369	52,149	175	562,011	1,788
Hs68 - 3							wt	21	0	0	52,724	119	832,166	1,856
Hs68 - 4							wt	1,647	0	0	104,514	40	1,087,819	3,110
Hs68 - 5							wt	0	0	47	6,341	425	271,530	521
Hs68 - 6							wt	367	0	0	25,147	0	491,608	246
Hs68 - 7							wt	0	16	0	30,704	0	488,513	192
Hs68 - 8							wt	0	0	0	7,682	0	604,049	344
MCF-7 - 1							wt	1,092	0	141	26,075	0	662,630	341
MCF-7 - 2							wt	1,349	0	341	17,550	0	632,258	621
MCF-7 - 3							wt	275	6	0	20,701	0	681,141	1,174
MCF-7 - 4							wt	0	0	0	8,565	237	484,097	737
MCF-7 - 5							wt	0	0	0	20,587	388	795,802	1,446
MCF-7 - 6							wt	0	0	0	0	0	33,599	0
MCF-7 - 7							wt	560	0	0	2,415	460	76,817	535
MCF-7 - 8							wt	1,159	0	126	65,998	43	777,026	2,150
OVCAR-4 - 1							wt	0	479	0	48,716	0	3,789,726	5,436
OVCAR-4 - 2							wt	0	314	77	29,888	0	417,174	388
OVCAR-4 - 3							wt	0	118	0	10,373	0	477,283	117
OVCAR-4 - 4							wt	396	123	89	38,345	0	581,094	508
OVCAR-4 - 5							wt	462	56	0	41,745	0	600,108	457
OVCAR-4 - 6							wt	278	59	208	7,508	356	427,310	554
OVCAR-4 - 7							wt	0	0	121	88,152	0	471,191	924
OVCAR-4 - 8							wt	0	283	92	11,388	0	509,238	91
SF539 - 1							wt	0	0	0	0	0	0	0



Table - Tissue Array  
413406\_1.xls

Tissue type	Release	tag	pos	NCI/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
U2OS-2	R2-384-1	sym	E01	487				SR-30							97			
U2OS-3	R2-384-1	sym	E03	488													99	
U2OS-4	R2-384-1	sym	E05	489													101	101
U2OS-5	R2-384-1	sym	E07	490													103	
U2OS-6	R2-384-1	sym	E09	491														
U2OS-7	R2-384-1	sym	E11	492														105
U2OS-8	R2-384-1	sym	E13	493														
A549-1	R2-384-1	sym	A01	333	CC-01	DD-01	MC-44	SR-01								107		
A549-2	R2-384-1	sym	A03	334			MC-45	SR-02								109		
A549-3	R2-384-1	sym	A05	335	CC-02										53			
A549-4	R2-384-1	sym	A07	336	CC-03	DD-02											5	
A549-5	R2-384-1	sym	A09	337	CC-04												7	7
A549-6	R2-384-1	sym	A11	338													9	
A549-7	R2-384-1	sym	A13	339	CC-05	DD-03												11
A549-8	R2-384-1	sym	A15	340			MC-02									13	13	
HCT-116-1	R2-384-1	sym	A17	341			MC-03									15		
HCT-116-2	R2-384-1	sym	A19	342			MC-04	SR-05							23	23		23
HCT-116-3	R2-384-1	sym	A21	343				SR-06							192			
HCT-116-4	R2-384-1	sym	A23	344														
HCT-116-5	R2-384-1	sym	A25	345														
HCT-116-6	R2-384-1	sym	A27	346														
HCT-116-7	R2-384-1	sym	A29	347														
HCT-116-8	R2-384-1	sym	A31	348														
HCT-116-9	R2-384-1	sym	A33	349														
HCT-116-10	R2-384-1	sym	A35	350														
HCT-116-11	R2-384-1	sym	A37	351														
HCT-116-12	R2-384-1	sym	A39	352														
HCT-116-13	R2-384-1	sym	A41	353														
HCT-116-14	R2-384-1	sym	A43	354														
HCT-116-15	R2-384-1	sym	A45	355														
HCT-116-16	R2-384-1	sym	A47	356														
HCT-116-17	R2-384-1	sym	A49	357														
HCT-116-18	R2-384-1	sym	A51	358														
HCT-116-19	R2-384-1	sym	A53	359														
HCT-116-20	R2-384-1	sym	A55	360														
HCT-116-21	R2-384-1	sym	A57	361														
HCT-116-22	R2-384-1	sym	A59	362														
HCT-116-23	R2-384-1	sym	A61	363														
HCT-116-24	R2-384-1	sym	A63	364														
HCT-116-25	R2-384-1	sym	A65	365														
HCT-116-26	R2-384-1	sym	A67	366														
HCT-116-27	R2-384-1	sym	A69	367														
HCT-116-28	R2-384-1	sym	A71	368														
HCT-116-29	R2-384-1	sym	A73	369														
HCT-116-30	R2-384-1	sym	A75	370														
HCT-116-31	R2-384-1	sym	A77	371														
HCT-116-32	R2-384-1	sym	A79	372														
HCT-116-33	R2-384-1	sym	A81	373														
HCT-116-34	R2-384-1	sym	A83	374														
HCT-116-35	R2-384-1	sym	A85	375														
HCT-116-36	R2-384-1	sym	A87	376														
HCT-116-37	R2-384-1	sym	A89	377														
HCT-116-38	R2-384-1	sym	A91	378														
HCT-116-39	R2-384-1	sym	A93	379														
HCT-116-40	R2-384-1	sym	A95	380														
HCT-116-41	R2-384-1	sym	A97	381														
HCT-116-42	R2-384-1	sym	A99	382														
HCT-116-43	R2-384-1	sym	A101	383														
HCT-116-44	R2-384-1	sym	A103	384														
HCT-116-45	R2-384-1	sym	A105	385														
HCT-116-46	R2-384-1	sym	A107	386														
HCT-116-47	R2-384-1	sym	A109	387														
HCT-116-48	R2-384-1	sym	A111	388														
HCT-116-49	R2-384-1	sym	A113	389														
HCT-116-50	R2-384-1	sym	A115	390														
HCT-116-51	R2-384-1	sym	A117	391														
HCT-116-52	R2-384-1	sym	A119	392														
HCT-116-53	R2-384-1	sym	A121	393														
HCT-116-54	R2-384-1	sym	A123	394														
HCT-116-55	R2-384-1	sym	A125	395														
HCT-116-56	R2-384-1	sym	A127	396														
HCT-116-57	R2-384-1	sym	A129	397														
HCT-116-58	R2-384-1	sym	A131	398														
HCT-116-59	R2-384-1	sym	A133	399														
HCT-116-60	R2-384-1	sym	A135	400														
HCT-116-61	R2-384-1	sym	A137	401														
HCT-116-62	R2-384-1	sym	A139	402														
HCT-116-63	R2-384-1	sym	A141	403														
HCT-116-64	R2-384-1	sym	A143	404														
HCT-116-65	R2-384-1	sym	A145	405														
HCT-116-66	R2-384-1	sym	A147	406														
HCT-116-67	R2-384-1	sym	A149	407														
HCT-116-68	R2-384-1	sym	A151	408														
HCT-116-69	R2-384-1	sym	A153	409														
HCT-116-70	R2-384-1	sym	A155	410														
HCT-116-71	R2-384-1	sym	A157	411														
HCT-116-72	R2-384-1	sym	A159	412														
HCT-116-73	R2-384-1	sym	A161	413														
HCT-116-74	R2-384-1	sym	A163	414														
HCT-116-75	R2-384-1	sym	A165	415														
HCT-116-76	R2-384-1	sym	A167	416														
HCT-116-77	R2-384-1	sym	A169	417														
HCT-116-78	R2-384-1	sym	A171	418														
HCT-116-79	R2-384-1	sym	A173	419														
HCT-116-80	R2-384-1	sym	A175	420														
HCT-116-81	R2-384-1	sym	A177	421														
HCT-116-82	R2-384-1	sym	A179	422														
HCT-116-83	R2-384-1	sym	A181	423														
HCT-116-84	R2-384-1	sym	A183	424														
HCT-116-85	R2-384-1	sym	A185	425														
HCT-116-86	R2-384-1	sym	A187	426														
HCT-116-87	R2-384-1	sym	A189	427														
HCT-116-88	R2-384-1	sym	A191	428														
HCT-116-89	R2-384-1	sym	A193	429														
HCT-116-90	R2-384-1	sym	A195	430														
HCT-116-91	R2-384-1	sym	A197	431														
HCT-116-92	R2-384-1	sym	A199	432														
HCT-116-93	R2-384-1	sym	A201	433														
HCT-116-94	R2-384-1	sym	A203	434														
HCT-116-95	R2-384-1	sym	A205	435														
HCT-116-96	R2-384-1	sym	A207	436														
HCT-116-97	R2-384-1	sym	A209	437														

Table - Tissue Array  
413406\_1.xls

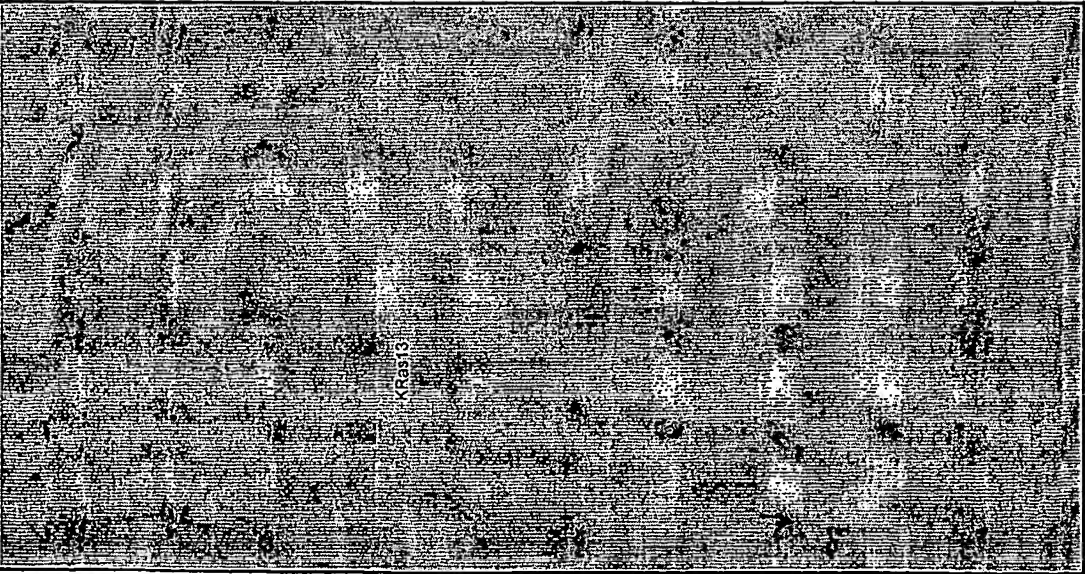
Ken Lip	Anglo	NCI-Ras	Hypoxia	PYK2	"Diabetes"
					

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
SF539 - 2							wt	277	145	0	9,827	11	651,624	427
SF539 - 3							wt	1,599	0	0	21,929	0	547,965	799
SF539 - 4							wt	0	179	0	19,159	0	449,192	668
SF539 - 5							wt	85	0	323	23,329	0	516,343	319
SF539 - 6							wt	0	108	0	7,791	0	662,214	668
SF539 - 7							wt	0	0	280	17,423	0	574,107	1,461
SF539 - 8							wt	0	58	0	38,279	0	550,442	1,574
WI 38 - 1							wt	391	90	25	54,921	0	498,076	672
WI 38 - 2							wt	725	671	339	93,773	270	339,104	2,179
WI 38 - 3							wt	1,101	524	65	40,084	0	414,523	945
WI 38 - 4							wt	6	752	0	37,054	0	620,563	661
WI 38 - 5							wt	0	289	0	7,149	0	552,397	1,748
WI 38 - 6							wt	161	0	0	10,543	0	559,003	724
WI 38 - 7							wt	385	63	92	31,968	0	832,476	1,298
WI 38 - 8							wt	80	157	0	5,957	0	1,071,262	1,559
HELA-2h-031899				79				533	0	0	660	127	61,571	1,378
HELA-4h-031899				81				0	0	0	1,627	0	98,196	929
HELA-8h-031899				83				0	232	0	174	0	91,313	810
HELA-0h-031899				86				0	0	0	293	0	41,904	164
HELA-6h-031899				88				305	0	0	462	0	56,525	291
HELA-8h-031899				90				0	0	0	0	0	80,592	531
HELA-10h-031899				92				0	0	0	2,217	0	50,148	0
HELA-11h-031899				94				0	0	0	0	0	45,767	0
HELA-12h-031899				96				496	0	0	231	0	43,735	297
NCH332M				146				0	138	0	1,805	0	69,902	1,926
NCI-H460				148				528	0	0	0	432	41,427	0
NCH522				150				0	101	0	0	960	64,259	1,627
SNB-19				152				515	0	0	0	217	15,805	214
SNB-75				154				10	0	0	1,207	0	36,729	0
SF-268				156				405	0	0	0	0	11,197	354
SF-295				158				612	630	0	760	532	43,210	1,364
CORF-CEM				160				0	0	0	599	0	2,830	40
DU-145				162				148	0	0	0	0	55,584	970
HCT 116				164				876	0	0	171	604	48,404	640
Csk-1				166				1,789	63	0	195	357	9,678	0
786-0				168				847	114	0	558	0	5,791	48
T-47D				169				404	55	0	1,659	0	13,446	531
CRL1441 RNA 8/30				181				1,738	0	0	0	590	4,867	0
781T untreated + DNase				183				1,066	16	0	0	52	4,220	168
KB poly A+				194				452	357	0	0	219	43,402	0
HOS poly A+				196				0	0	0	0	904	58,885	468
ACHN				198				426	236	0	193	3,740	67,534	748
UACC-62				200				0	282	0	459	7	12,401	0
MCF-7/ADR-RES				202				776	0	0	209	0	44,167	0
UTOS (Mundy) poly a+				204				0	143	0	373	359	13,394	115
WISH (Collagen) poly A+				206				1,552	0	0	0	0	3,311	0
458 medullo mRNA				208				0	290	0	4,761	0	38,074	0



Table - Tissue Array  
413406\_1.xls

Tissue type	Release	tag	pos	NCI/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
SF539-2	R2384-1	57m	B07	471				SR-10					31		131			
SF539-3	R2384-1	57m	B09	472									33					
SF539-4	R2384-1	57m	B11	473									35					
SF539-5	R2384-1	57m	B13	474		DD-14							37					
SF539-6	R2384-1	57m	B15	475		DD-15							39					
SF539-7	R2384-1	57m	B17	476			MC-14						41					
SF539-8	R2384-1	57m	B19	477			MC-15						43					
WI 38-1	R2384-1	57m	E08	494	CC-21	DD-31	MC-29	SR-33						104				
WI 38-2	R2384-1	57m	E10	495				SR-34						106				
WI 38-3	R2384-1	57m	E12	496	CC-22													
WI 38-4	R2384-1	57m	E14	497	CC-23	DD-32												
WI 38-5	R2384-1	57m	E16	498	CC-24													
WI 38-6	R2384-1	57m	E18	499		DD-33												
WI 38-7	R2384-1	57m	E20	500	CC-25		MC-50											
WI 38-8	R2384-1	57m	E22	501			MC-51											
HELA-2H-031899	R1384-1	57m	D07	148														
HELA-4H-031899	R1384-1	57m	D09	149														
HELA-9H-031899	R1384-1	57m	D11	152														
HELA-10H-031899	R1384-1	57m	D14	157														
HELA-5H-031899	R1384-1	57m	D16	150														
HELA-8H-031899	R1384-1	57m	D18	151														
HELA-10H-031899	R1384-1	57m	D20	153														
HELA-11H-031899	R1384-1	57m	D22	154														
HELA-12H-031899	R1384-1	57m	D24	155														
NCI-H460	R1384-1	57m	G02	5														
NCI-H522	R1384-1	57m	G04	6														
NCI-H522	R1384-1	57m	G06	7														
SNB-19	R1384-1	57m	G08	16									152					
SNB-75	R1384-1	57m	G10	17														
SF-295	R1384-1	57m	G12	19									156					
COR-CEM	R1384-1	57m	G14	20									158					
DU-145	R1384-1	57m	G16	22									160					
HCT-116	R1384-1	57m	G18	28														
Cal-51	R1384-1	57m	G20	31														
7850	R1384-1	57m	G22	39														
1470	R1384-1	57m	G24	42														
CRJ-1441RNA-820	R1384-1	57m	H01	58														
7811 (untreated + DNase)	R1384-1	57m	H13	79														
KB poly A+	R1384-1	57m	H15	73														
KB poly A+	R1384-1	57m	I02	59														
KB poly A+	R1384-1	57m	I04	63														
ACHN	R1384-1	57m	I06	41														
ACHN	R1384-1	57m	I08	49														
UACC-25	R1384-1	57m	I10	53														
MC7	R1384-1	57m	I12	65														
UTOS (Mundy) poly A+	R1384-1	57m	I14	66														
WISH (Collagen) poly A+	R1384-1	57m	I16	69														
453 medium RNA	R1384-1	57m																

Table - Tissue Array  
413406\_1.xls

Ken Lip	Anglo	NCI-Ras	Hypoxia	PYK2	"Diabetes"



Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
CCL137 RNA 3/21/88				218				0	0	0	392	65	35,013	425
WI-38 72h 0.5%FBS, 24h 10% FBS				219				0	0	0	0	0	43,199	0
CRL1441 + TPA (24h) 8/30				220				0	56	0	366	0	13,416	59
HOP-82				241				0	23	0	0	0	22,647	186
MOLT-4				242				0	92	0	309	26	21,250	434
EKVX				243				0	250	0	960	0	77,320	2,875
HL-60				244				0	0	0	908	480	24,392	0
NCI-H23				245				1,129	0	0	211	0	93,283	2,482
RPMI 8228				246				0	0	0	799	0	87,784	1,931
A549/ATCC				247				226	0	0	0	211	5,668	0
SR				248				329	0	0	0	0	15,586	298
OVCAR-3				249				112	0	0	0	594	15,287	240
HCT-15				250				284	387	0	2,254	94	30,811	211
OVCAR-4				251				0	278	0	0	0	4,121	166
UO-31				252				0	0	0	0	30	7,831	96
OVCAR-5				253				181	265	0	1,579	756	28,038	1,658
SN12C				254				0	151	0	0	0	5,978	0
OVCAR-8				255				0	7	0	952	591	67,456	1,954
LOX IMVI				256				239	624	0	2,845	954	25,761	1,176
IGROV1				257				0	0	0	1,111	3	56,485	1,865
SK-MEL-2				258				0	0	0	0	0	2,107	619
SK-OV-3				259				350	0	0	0	0	3,010	85
SK-MEL-5				260				0	0	0	0	0	1,573	0
SF-539				261				728	0	0	859	0	9,243	202
SK-MEL-28				262				148	0	0	1,157	273	23,788	634
K-562				263				313	683	0	783	597	4,477	300
UACC-257				264				245	0	0	26	0	2,255	330
M14				265				90	242	0	0	214	129,614	1,060
MCF7				267				204	0	0	16,366	0	432,526	111
MDA-MB-435				269				90	0	0	1,131	0	14,733	0
HT279				270				0	323	0	0	0	19,273	86
MDA-N				271				12	176	0	89	0	9,135	317
Y78 poly A+				273				463	107	0	0	20	89,877	0
KHOS poly A+				289				0	0	0	0	0	75,507	0
HTB38 24h TPA RNA 6/23				300				5	50	0	0	86	34,541	0
HELA-EXP-031899				313				0	0	0	110	0	55,070	45
HTB36 0h RNA				322				0	20	0	0	0	49,195	54
HT347				323				0	282	0	521	0	46,274	127
458 medullo RNA				324				0	0	0	0	0	26,987	60
NCI-H226				336				214	352	0	0	257	2,982	0
HOP-82				337				219	0	0	0	0	11,386	157
MDA-MB-231				338				0	0	0	577	277	51,134	901
U251				339				0	237	0	4,549	515	55,589	372
PT cells poly A+				340				130	106	0	0	261	8,973	0
PC-3				341				0	284	0	0	1,140	65,836	1,366
HCC-2998				343				803	0	0	375	639	118,307	958
SW-620				345				0	0	79	22,607	0	467,592	1,300

Table - Tissue Array  
413406\_1.xls

Tissue type	Release	tag	pos	NCI/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
CO-137 RNA 3/21/89	R13841	asym	07	77														
WI-3972h0.5%FBS 24h10% FBS	R13841	asym	03	96														
CR-1441 TPA (24h) 850	R13841	asym	04	78														
HOP-92	R13841	asym	K01	11														
MOLT-4	R13841	asym	K02	24														
EVX-45	R13841	asym	K03	21										242				
HL-60	R13841	asym	K04	25														
NCH23	R13841	asym	K05	3														
RPMI 8226	R13841	asym	K06	26														
A549/ATCC	R13841	asym	K07	8														
SR	R13841	asym	K08	27														
OVCAR-3	R13841	asym	K09	10														
HCT15	R13841	asym	K10	34														
OVCAR-4	R13841	asym	K11	11														
UO-31	R13841	asym	K12	36														
OVCAR-5	R13841	asym	K13	12														
SW-62	R13841	asym	K14	37														
OVCAR-6	R13841	asym	K15	13														
LOX IMV	R13841	asym	K16	24														
GRV-1	R13841	asym	K17	14														
SK-MEL-2	R13841	asym	K18	46														
SK-OV-3	R13841	asym	K19	15														
SK-MEL-5	R13841	asym	K20	47														
SF-539	R13841	asym	K21	21														
SK-MEL-28	R13841	asym	K22	48														
K-562	R13841	asym	K23	23														
UACC257	R13841	asym	K24	50														
M14	R13841	asym	L01	51														
MCF-7	R13841	asym	L03	52														
MDA-MB-435	R13841	asym	L05	53														
HT29	R13841	asym	L06	185														
MDA-N	R13841	asym	L07	57														
Y79	R13841	asym	L09	61														
KHOS (poly A)	R13841	asym	L01	63														
HT69 24h TPA RNA 8/23	R13841	asym	M12	82														
HELA EXP 031899	R13841	asym	N01	156														
HTB39 OR RNA	R13841	asym	N10	71														
HT 347	R13841	asym	N11	200														
458 medulo RNA	R13841	asym	N12	74														
NCH-228	R13841	asym	N24	4														
HOP-92	R13841	asym	O01	9														
MDA-MB-231	R13841	asym	O02	55														
U251	R13841	asym	O03	18														
PT cells poly A	R13841	asym	O04	60														
PC-3	R13841	asym	O05	29														
HCC-2998	R13841	asym	O07	10														
SW480	R13841	asym	O09	32														

**Table - Tissue Array**  
**413406\_1.xls**

[illegible]

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK396 ID#NA_43
HT192				345				0	0	0	0	0	74,830	83
COLO 205				347				621	6	0	0	0	10,126	0
HT218				348				0	0	0	0	0	20,377	0
KM-12				349				860	259	0	623	0	21,590	0
HT151				350				0	411	0	347	0	60,081	414
A498				351				0	428	0	0	378	9,226	0
HT393				352				0	0	0	694	0	57,432	0
RXF 393				353				0	0	0	0	0	6,338	82
TK-10				355				191	0	0	2,314	0	14,179	0
Malme-3M				357				0	0	0	1,531	885	10,560	174
Hs 578T				359				212	0	0	0	0	2,568	0
HT213			50					886	0	0	0	0	48,907	0
HT288			52					0	224	0	693	0	74,176	18
HT139			54					0	0	0	631	0	67,320	102
HT155			56					0	0	0	30	0	61,739	0
HT163			58					48	0	0	0	345	80,453	275
HT170			60					0	184	0	773	0	63,967	0
HT172			62					0	0	0	0	0	48,537	0
HT138			63					10	1	0	741	0	46,416	156
HT178			64					0	14	0	301	0	57,882	330
HT154			65					0	166	0	0	0	69,411	89
HT180			66					47	57	0	0	0	53,278	214
HT169			67					0	0	0	0	0	23,349	0
HT189			68					182	0	0	0	0	63,765	243
HT143			69					0	0	0	86	0	43,069	129
HT190			70					0	0	0	10	0	51,091	22
HT145			71					372	0	0	0	0	58,721	0
HT227			72					101	0	0	173	29	33,093	0
HT302			73					56	207	0	0	0	69,267	275
HT314			74					0	20	0	1,141	0	60,939	287
HT317			76					0	0	0	0	0	51,326	484
Medulloblastoma #425 11/8			77					0	74	0	774	0	111,765	60
HT323			78					0	0	0	0	0	76,589	339
HT327			80					0	251	0	691	432	62,791	338
HT335			82					0	25	0	176	0	69,031	0
HT146			85					227	0	0	0	405	7,160	0
HT348			87					359	0	0	0	0	69,239	1,165
HT311			170					0	75	0	0	0	31,571	248
HT398			185					3,484	281	0	0	259	5,903	31
HT140			187					662	0	0	0	0	1,081	0
HT281			189					549	0	0	0	231	5,431	0
HT372			191					236	0	0	0	0	11,267	0
TCGP			207					0	0	0	0	265	9,599	16
HT160			216					0	0	0	0	0	2,822	0
HT307			217					144	149	0	0	0	27,812	84
HT369			224					0	103	0	0	0	45,642	0
HT370			226					0	130	0	0	0	33,703	22



Table - Tissue Array  
413406\_1.xls

Tissue_type	Release	tag	pos	NC/lec	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
HT192	R13841	asym	O10	179														
COLO205	R13821	asym	O11	13														
HT218	R13841	asym	O12	182														
KM12	R13841	asym	O13	35														
HT151	R13841	asym	O14	164														
A498	R13841	asym	O15	538														
HT393	R13841	asym	O16	215														
RA1393	R13841	asym	O17	49														
TK10	R13841	asym	O19	43														
Worm	R13841	asym	O21	45														
HT213	R13841	asym	O21	54														
HT288	R13841	asym	O21	180														
HT399	R13841	asym	O24	187														
HT109	R13841	asym	O26	158														
HT163	R13841	asym	O28	169														
HT170	R13841	asym	O28	171														
HT172	R13841	asym	O28	173														
HT138	R13841	asym	O28	174														
HT178	R13841	asym	O28	157														
HT154	R13841	asym	O28	173														
HT180	R13841	asym	O28	165														
HT180	R13841	asym	O28	176														
HT180	R13841	asym	O28	172														
HT180	R13841	asym	O28	177														
HT180	R13841	asym	O28	160														
HT180	R13841	asym	O28	178														
HT180	R13841	asym	O28	161														
HT180	R13841	asym	O28	184														
HT180	R13841	asym	O28	183														
HT180	R13841	asym	O28	193														
HT180	R13841	asym	O28	194														
HT180	R13841	asym	O28	83														
HT180	R13841	asym	O28	195														
HT180	R13841	asym	O28	196														
HT180	R13841	asym	O28	198														
HT180	R13841	asym	O28	162														
HT180	R13841	asym	O28	201														
HT180	R13841	asym	O28	191														
HT180	R13841	asym	O28	159														
HT180	R13841	asym	O28	184														
HT180	R13841	asym	O28	206														
HT180	R13841	asym	O28	176														
HT180	R13841	asym	O28	189														
HT180	R13841	asym	O28	203														
HT180	R13841	asym	O28	204														
HT180	R13841	asym	O28	204														

Table - Tissue Array  
413406\_1.xls

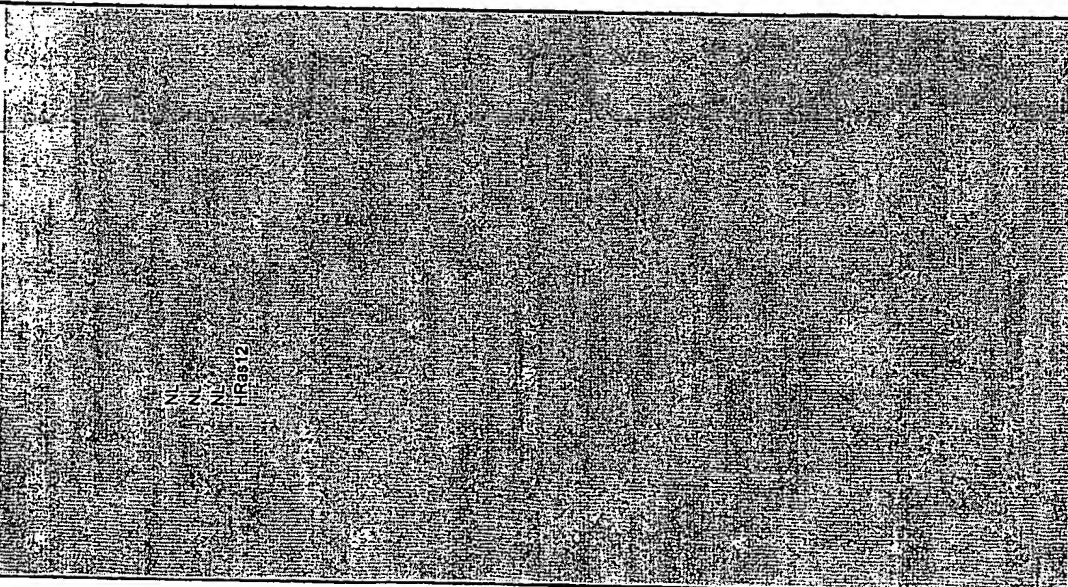
Ken Lip	Anglo	NCI-Ras	Hypoxia	PYK2	"Diabetes"
					

Table - Tissue Array  
413406\_1.xls

Tissue	Tumor sym	Normal sym	Tumor 1o	Tumor cells	Normal	Endos	p53	SGK187 ID#NA_1	SGK124 ID#NA_9	SGK386 ID#NA_21	SGK003 ID#NA_22	SGK093 ID#NA_31	SGK074 ID#NA_32	SGK398 ID#NA_43
HT371			228					0	56	0	576	0	46,710	453
HT377			230					0	132	0	0	0	35,666	354
HT382			236					101	328	0	0	0	28,197	45
Her2/neu RNA			281					1,184	396	0	2,805	0	0	0
HT334			299					0	0	0	0	0	111,843	377
HT338			301					0	56	0	0	0	28,893	0
HT392			315					460	0	0	0	0	19,458	155
HT394			317					0	0	0	46	62	37,089	235
HT312			318					181	36	0	0	245	28,897	0
HT182			325					0	0	0	0	0	17,011	0
HT385			358					233	0	0	0	0	7,637	39
HT197			360					0	383	0	0	0	13,226	0
458 medullio RNA								138	154	0	183	0	90,329	690
CRL1572 3/17/89								0	0	0	248	0	56,547	0
h fibroblasts 3/31/92 #12								423	390	0	0	102	12,530	184
h keratinocytes 2/25/92 #10								64	0	0	0	419	15,538	0
HT308								899	0	0	0	0	54,950	21
HT388								0	31	0	0	80	86,952	0
HT378								0	0	0	-40	0	56,777	707
HT385								0	0	0	846	0	33,000	0

Table - Tissue Array  
413406\_1.xls

Tissue type	Release	tag	pos	NCL/etc	CC	DD	MC	SR	ANG	PRO	Mets	Prostate	Neuro	Heme	Serum	Mitosis	Cell Cycle	DNA Dam
HT371	R1384	15	112	205														
HT377	R1384	15	112	205														
HT382	R1384	15	112	205														
neuroblastoma RNA	R1384	15	112	205														
HT334	R1384	15	112	205														
HT335	R1384	15	112	205														
HT392	R1384	15	112	205														
HT393	R1384	15	112	205														
HT394	R1384	15	112	205														
HT395	R1384	15	112	205														
HT396	R1384	15	112	205														
HT397	R1384	15	112	205														
HT398	R1384	15	112	205														
HT399	R1384	15	112	205														
HT400	R1384	15	112	205														
HT401	R1384	15	112	205														
HT402	R1384	15	112	205														
HT403	R1384	15	112	205														
HT404	R1384	15	112	205														
HT405	R1384	15	112	205														
HT406	R1384	15	112	205														
HT407	R1384	15	112	205														
HT408	R1384	15	112	205														
HT409	R1384	15	112	205														
HT410	R1384	15	112	205														
HT411	R1384	15	112	205														
HT412	R1384	15	112	205														
HT413	R1384	15	112	205														
HT414	R1384	15	112	205														
HT415	R1384	15	112	205														
HT416	R1384	15	112	205														
HT417	R1384	15	112	205														
HT418	R1384	15	112	205														
HT419	R1384	15	112	205														
HT420	R1384	15	112	205														
HT421	R1384	15	112	205														
HT422	R1384	15	112	205														
HT423	R1384	15	112	205														
HT424	R1384	15	112	205														
HT425	R1384	15	112	205														
HT426	R1384	15	112	205														
HT427	R1384	15	112	205														
HT428	R1384	15	112	205														
HT429	R1384	15	112	205														
HT430	R1384	15	112	205														
HT431	R1384	15	112	205														
HT432	R1384	15	112	205														
HT433	R1384	15	112	205														
HT434	R1384	15	112	205														
HT435	R1384	15	112	205														
HT436	R1384	15	112	205														
HT437	R1384	15	112	205														
HT438	R1384	15	112	205														
HT439	R1384	15	112	205														
HT440	R1384	15	112	205														
HT441	R1384	15	112	205														
HT442	R1384	15	112	205														
HT443	R1384	15	112	205														
HT444	R1384	15	112	205														
HT445	R1384	15	112	205														
HT446	R1384	15	112	205														
HT447	R1384	15	112	205														
HT448	R1384	15	112	205														
HT449	R1384	15	112	205														
HT450	R1384	15	112	205														
HT451	R1384	15	112	205														
HT452	R1384	15	112	205														
HT453	R1384	15	112	205														
HT454	R1384	15	112	205														
HT455	R1384	15	112	205														
HT456	R1384	15	112	205														
HT457	R1384	15	112	205														
HT458	R1384	15	112	205														
HT459	R1384	15	112	205														
HT460	R1384	15	112	205														
HT461	R1384	15	112	205														
HT462	R1384	15	112	205														
HT463	R1384	15	112	205														
HT464	R1384	15	112	205														
HT465	R1384	15	112	205														
HT466	R1384	15	112	205														
HT467	R1384	15	112	205														
HT468	R1384	15	112	205														
HT469	R1384	15	112	205														
HT470	R1384	15	112	205														
HT471	R1384	15	112	205														
HT472	R1384	15	112	205														
HT473	R1384	15	112	205														
HT474	R1384	15	112	205														
HT475	R1384	15	112	205														
HT476	R1384	15	112	205														
HT477	R1384	15	112	205														
HT478	R1384	15	112	205														
HT479	R1384	15	112	205														
HT480	R1384	15	112	205														
HT481	R1384	15	112	205														
HT482	R1384	15	112	205														
HT483	R1384	15	112	205														
HT484	R1384	15	112	205														
HT485	R1384	15	112	205														
HT486	R1384	15	112	205														
HT487	R1384	15	112	205														
HT488	R1384	15	112	205														
HT489	R1384	15	112	205														
HT490	R1384	15	112	205														
HT491	R1384	15	112	205														
HT492	R1384	15	112	205														
HT493	R1384	15	112	205														
HT494	R1384	15	112	205														
HT495	R1384	15	112	205														
HT496	R1384	15	112	205														
HT497	R1384	15	112	205														
HT498	R1384	15	112	205														
HT499	R1384	15	112	205														
HT500	R1384	15	112	205														



Table - Tissue Array  
413406\_1.xls

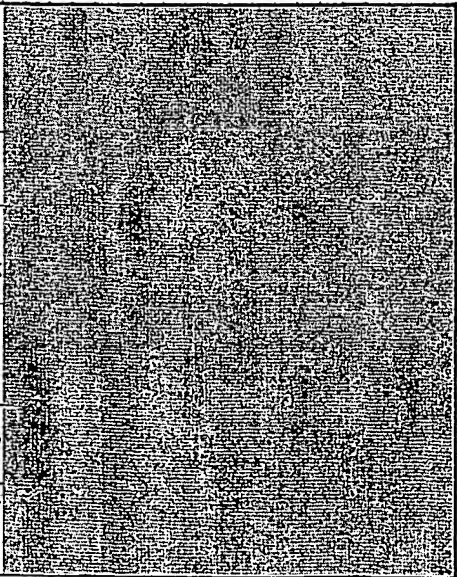
Ken Lip	Angio	NCI-Ras	Hypoxia	PYK2	"Diabetes"
					

Table 7 gives results of a PCR screen of 48 human cDNA sources for 26 of the kinases represented in this application. A plus sign (+) indicates the presence of a band on an agarose gel of the expected size for the target kinase. A negative sign (-) indicates that the PCR product of the expected size was absent. The genes represented on this table are: SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:56.

Table - PCR Expression Analysis

Tissue name	gel-well	96-well pos	RNA source	Tissue/Cells	SGK069	SGK110	SGK033	SGK254	SGK411	SGK027	SGK046b	SGK089	SGK003	SGK066	SGK041	SGK038	SGK429
1	A01	Clontech	Normal tissue	NA #6	NA #7	NA #8	NA #10	NA #12	NA #13	NA #14	NA #16	NA #22	NA #23	NA #26	NA #26	NA #26	NA #28
2	A02	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	A03	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	A04	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	A05	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	A06	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	A07	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	A08	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	A09	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	A10	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	A11	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	A12	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	B01	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	B02	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	B03	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	B04	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	B05	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	B06	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19	B07	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	B08	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	B09	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	B10	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	B11	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
24	B12	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	C01	Sugen	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	C02	Sugen	Normal Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27	C03	Sugen	Normal Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
28	C04	Sugen	Normal Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	C05	Sugen	Normal Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	C06	Sugen	Tumor	-	-	-	-	-	-	-	-	-	-	-	-	-	-
31	C07	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
32	C08	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33	C09	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
34	C10	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
35	C11	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
36	C12	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
37	D01	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
38	D02	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
39	D03	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	D04	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
41	D05	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
42	D06	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
43	D07	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
44	D08	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
45	D09	GP	Tumor	-	-	-	-	-	-	-	-	-	-	-	-	-	-
46	D10	GP	Normal Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
47	D11	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-
48	D12	NCI	Cancer Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table - PCR Expression Analysis

Tissue name	gel-well	96-well pos	RNA_source	Tissue/Cells	SGK093	SGK087	SGK009	SGK421	SGK396	SGK037	SGK060	SGK080	SGK002	SGK058	SGK035	SGK075	SGK007
fetal liver - h	1	A01	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	+	-	-
thymus - h	2	A02	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	-	-
pancreas - h	3	A03	Clontech	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	+	-
pituitary gland - h	4	A04	Clontech	Normal tissue	-	-	-	-	-	-	+	-	-	-	+	-	-
placenta - h	5	A05	Clontech	Normal tissue	+	-	-	-	-	-	-	-	-	+	+	+	-
prostate, h	6	A06	Clontech	Normal tissue	+	-	-	-	-	-	+	-	-	+	+	-	-
salivary gl. - h	7	A07	Clontech	Normal tissue	+	+	+	-	-	-	+	-	-	-	-	-	-
skeletal muscle - h	8	A08	Clontech	Normal tissue	+	-	-	-	-	-	+	-	-	-	+	-	-
small intestine - h	9	A09	Clontech	Normal tissue	+	+	-	-	-	+	-	-	-	+	+	-	-
spinal cord - h	10	A10	Clontech	Normal tissue	+	+	-	-	-	-	-	-	-	-	-	-	-
Spleen - h	11	A11	Clontech	Normal tissue	+	+	+	-	-	-	-	-	-	-	-	-	-
stomach - h	12	A12	Clontech	Normal tissue	+	-	+	-	-	-	+	-	-	-	+	-	-
thyroid gland - h	13	B01	Clontech	Normal tissue	+	+	+	-	-	-	+	-	-	-	+	-	-
trachea - h	14	B02	Clontech	Normal tissue	+	-	+	-	-	+	+	-	-	+	-	-	-
uterus - h	15	B03	Clontech	Normal tissue	+	-	-	-	+	-	-	-	-	-	-	-	-
adrenal gland - h	16	B04	Clontech	Normal tissue	+	+	-	-	-	-	-	-	-	-	-	-	-
adrenal gland - h	17	B05	Clontech	Normal tissue	+	+	-	-	-	-	-	-	-	-	-	-	-
fetal brain - h	18	B06	Clontech	Normal tissue	+	+	+	-	-	-	-	-	-	+	+	-	-
fetal kidney - b	19	B07	Clontech	Normal tissue	+	+	+	-	-	+	+	-	-	+	+	-	-
fetal lung - h	20	B08	Clontech	Normal tissue	+	-	-	-	-	-	-	-	-	-	+	-	-
heart - h	21	B09	Clontech	Normal tissue	+	-	+	-	-	-	-	-	-	-	+	-	-
kidney - h	22	B10	Clontech	Normal tissue	+	-	+	-	-	-	-	-	-	-	+	-	-
liver - h	23	B11	Clontech	Normal tissue	+	+	+	-	-	+	+	-	-	-	+	-	-
lung - h	24	B12	Clontech	Normal tissue	+	-	-	-	-	-	+	-	-	-	-	-	-
lymph node - h	25	C01	Sugen	Normal tissue	-	-	-	-	-	-	-	-	-	-	-	+	-
Heart - h	26	C02	Sugen	Normal Cell line	-	-	-	-	-	-	-	-	-	-	-	-	-
HPAEC - Renal proximal tubule epithelial cells	27	C03	Sugen	Normal Cell line	+	-	-	-	-	-	-	-	-	-	+	-	-
RPTEC - Mammary epithelial cells	28	C04	Sugen	Normal Cell line	+	-	-	-	-	-	-	-	-	-	+	-	-
HMEC - Coronary artery endothelial cells	29	C05	Sugen	Normal Cell line	+	-	-	-	-	-	-	-	-	-	+	-	-
HCAREC - Coronary artery endothelial cells	30	C06	Sugen	Normal Cell line	+	-	-	-	-	-	-	-	-	-	+	-	-
458 medullo RNA - Neuroblastoma	31	C07	NCI	Tumor	+	-	-	-	-	-	-	-	-	-	+	-	-
A549/ATCC - Lung carcinoma	32	C08	NCI	Cancer Cell line	+	-	-	-	-	-	+	-	-	-	+	+	-
MDA-MB-231 - Breast adenocarcinoma, pleural effusion	33	C09	NCI	Cancer Cell line	+	-	-	-	-	-	+	-	-	-	+	+	-
HS 578T - Ductal breast carcinoma	34	C10	NCI	Cancer Cell line	+	-	-	-	-	-	+	-	-	-	+	+	-
MCF-7/ADR-RES - Breast adenocarcinoma	35	C11	NCI	Cancer Cell line	+	-	+	-	-	-	-	-	-	-	+	+	-
Malme-3M - Malignant melanoma, metastasis to lung	36	C12	NCI	Cancer Cell line	+	-	-	-	-	-	-	-	-	-	+	+	-
A498 - Kidney carcinoma	37	D01	NCI	Cancer Cell line	+	+	+	-	-	-	-	-	-	-	+	+	-
COLO 205 - Colon adenocarcinoma	38	D02	NCI	Cancer Cell line	+	+	+	-	-	+	-	-	-	-	+	+	-
CCRF-CEM - ALL Acute lymphoblastic leukemia	39	D03	NCI	Cancer Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-
SF-539 - Glioblastoma	40	D04	NCI	Cancer Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-
SF-295 - Glioblastoma	41	D05	NCI	Cancer Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-
U251 - Glioblastoma	42	D06	NCI	Cancer Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-
SNB-19 - Glioblastoma	43	D07	NCI	Cancer Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-
OVCAR-4 - Ovary adenocarcinoma	44	D08	NCI	Cancer Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-
OVCAR-3 - Ovary adenocarcinoma	45	D09	GP	Tumor	+	+	+	-	-	-	+	-	-	-	+	+	-
TCGF - Testicular carcinoma	46	D10	GP	Normal Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-
HMEC - Coronary artery endothelial cells	47	D11	NCI	Cancer Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-
HOP-62 - Lung adenocarcinoma	48	D12	NCI	Cancer Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-
NCI-H522 - Lung adenocarcinoma				Cancer Cell line	+	+	+	-	-	-	+	-	-	-	+	+	-

Table 8 contains results of probing a multiple tissue expression blot from Clontech (<http://www.clontech.com/>) with two of the genes from this application, SGK093 (SEQ ID NO:31); and SGK138, (SEQ ID NO:53).

Table - Clontech Multiple Tissue Blot

Clontech Tissue Blot	SGK093 ID#NA 31	SGK188 ID#NA 33
accumbens nucleus	365	3092
adrenal gland	64	570
amygdala	309	4796
aorta	360	309
apex of the heart	305	517
appendix	677	1695
atrium, left	480	448
atrium, right	561	232
bladder	32	539
bone marrow	408	421
Burkitt's lymphoma, Daudi	399	0
Burkitt's lymphoma, Raji	430	206
caudate nucleus	187	8750
cerebellum left	541	1960
cerebellum, right	680	1436
cerebral cortex	126	6185
colon, ascending	3480	7453
colon, descending	11042	3658
colon, transverse	4419	4555
colorectal adeno-carcinoma, SW480	1379	1389
corpus callosum	142	2221
duodenum	4014	2200
esophagus	3973	612
fetal brain	321	5202
fetal heart	627	709
fetal kidney	12081	199
fetal liver	9488	424
fetal lung	448	334
fetal spleen	210	110
fetal thymus	225	361
frontal lobe	432	3891
heart	285	394
HeLa S3	500	390
hippocampus	652	6100
ileum	2598	2072
ileocecum	1307	3820
inter-ventricular septum	0	100
jujenum	2588	858
kidney	16981	162
leukemia, HL-60	879	242
leukemia, K562	14760	246
leukemia, MOLT-4	201	344
liver	494	649
lung	928	710
lung carcinoma, A549	815	1439
lymph node	1044	1503
mammary gland	1659	1138
medulla oblongata	360	3260
occipital lobe	199	7343
ovary	1056	275
pancreas	858	1696
paracentral gyrus cerebral cortex	195	4394
parietal lobe	219	6799
peripheral blood leukocyte	19	342
pituitary gland	2009	4355
placenta	284	543
pons	0	4160
prostate	10296	984
putamen	85	5033
rectum	11181	4398
salivary gland	281	1232
skeletal muscle	156	214
spinal cord	895	1084
spleen	124	603
stomach	1941	2124
substantia nigra	332	2418
temporal lobe	0	7253
testis	607	31121
thalamus	502	2902
thymus	182	127
thyroid gland	128	24
trachea	1393	926
uterus	420	443
ventricle, left	502	583
ventricle, right	268	859
whole brain	258	5291

## EXAMPLES

The examples below are not limiting and are merely representative of various aspects and features of the present invention. The examples below demonstrate the isolation and characterization of the nucleic acid molecules according to the invention, as well as the polypeptides they encode.

### **EXAMPLE 1: Identification and Characterization of Genomic Fragments Encoding Protein Kinases**

#### **Materials and Methods**

Novel kinases were identified from the Celera human genomic sequence databases, and from the public Human Genome Sequencing project (<http://www.ncbi.nlm.nih.gov/>) using a hidden Markov model (HMMR) built with 70 mammalian and yeast kinase catalytic domain sequences. These sequences were chosen from a comprehensive collection of kinases such that no two sequences had more than 50% sequence identity. The genomic database entries were translated in six open reading frames and searched against the model using a Timelogic Decypher box with a Field programmable array (FPGA) accelerated version of HMMR2.1. The DNA sequences encoding the predicted protein sequences aligning to the HMMR profile were extracted from the original genomic database. The nucleic acid sequences were then clustered using the Pangea Clustering tool to eliminate repetitive entries. The putative protein kinase sequences were then sequentially run through a series of queries and filters to identify novel protein kinase sequences. Specifically, the HMMR identified sequences were searched using BLASTN and BLASTX against a nucleotide and amino acid repository containing 634 known human protein kinases and all subsequent new protein kinase sequences as they are identified. The output was parsed into a spreadsheet to facilitate elimination of known genes by manual inspection. Two models were developed, a "complete" model and a "partial" or Smith Waterman model. The partial model was used to identify sub-catalytic kinase domains, whereas the complete model was used to identify complete catalytic domains. The selected hits were then queried using BLASTN against the public nma and EST databases to confirm they are indeed

unique. In some cases the novel genes were judged to be homologues of previously identified rodent or vertebrate protein kinases.

Extension of partial DNA sequences to encompass the full-length open-reading frame was carried out by several methods. Iterative blastn searching of the cDNA databases listed in Table 9 was used to find cDNAs that extended the genomic sequences. "LifeSeqGold" databases are from Incyte Genomics, Inc (<http://www.incyte.com/>). NCBI databases are from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). All blastn searches were conducted using a penalty for a nucleotide mismatch of -3 and reward for a nucleotide match of 1. The gapped blast algorithm is described in: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402).

Extension of partial DNA sequences to encompass the full-length open-reading frame was also carried out by iterative searches of genomic databases. The first method made use of the Smith-Waterman algorithm to carry out protein-protein searches of a close protein homologue to the partial. The target databases consisted of Genscan and open-reading frame (ORF) predictions of all human genomic sequence derived from the human genome project (HGP) as well as from Celera. The complete set of genomic databases searched is shown in Table 10, below. Genomic sequences encoding potential extensions were further assessed by blastx analysis against the NCBI nonredundant database to confirm the novelty of the hit. The extending genomic sequences were incorporated into the cDNA sequence after removal of potential introns using the Seqman program from DNASTar. The default parameters used for Smith-Waterman searches were as shown next. Matrix: blosum 62; gap-opening penalty: 12; gap extension penalty: 2. Genscan predictions were made using the Genscan program as detailed in Chris Burge and Sam Karlin "Prediction of Complete Gene Structures in Human Genomic DNA", JMB (1997) 268(1):78-94). ORF predictions from genomic DNA were made using a standard 6-frame translation.

Another method for defining DNA extensions from genomic sequence used iterative searches of genomic databases through the Genscan program to predict exon splicing. These



predicted genes were then assessed to see if they represented "real" extensions of the partial genes based on homology to related kinases.

- Another method involved using the Genewise program  
(<http://www.sanger.ac.uk/Software/Wise2/>) to predict potential ORFs based on homology or
- 5 HMM to the closest homologue. Genewise requires two inputs, the homologous protein, and genomic DNA containing the gene of interest. The genomic DNA was identified by blastn searches of Celera and Human Genome Project databases. The homologues were identified by blastp searches of the NCBI non-redundant protein database (NR) with the predicted protein sequence derived from the HMM search of the genomic database. Genewise compares the
- 10 protein sequence to a genomic DNA sequence, allowing for introns and frameshifting errors.

TABLE 9.

Databases used for cDNA-based sequence extensions

Database	Database Date
LifeGold templates	Oct 2000
LifeGold compseqs	Oct 2000
LifeGold compseqs	Oct 2000
LifeGold compseqs	Oct 2000
LifeGold fl	Oct 2000
LifeGold flt	Oct 2000
NCBI human Ests	Oct 2000
NCBI murine Ests	Oct 2000
NCBI nonredundant	Oct 2000

Database	Number of entries	Database Date
----------	-------------------	---------------

**TABLE 10.**  
Databases used for genomic-based sequence extensions

Database	Number of entries	Database Date
Celera v. 1-5	5,306,158	Jan 19/00
Celera v. 6-10	4,209,980	Mar24/00
Celera v. 11-14	7,222,425	Apr 24/00
Celera v. 15	243,044	May14/00
Celera v. 16-17	25,885	Apr 04/00
Celera Assembly 5 (R1.25)	3,313	Oct 13/00
Celera Assembly 4 (R1.24)	636,234	Aug 28/00
Celera Assembly 3 (R 1.22, 1.23)	1,132,900	Jul 21/00
HGP Phase 0	4,944	May 04/00
HGP Phase 1	28,478	May05/00
HGP Phase 2	1,508	May04/00
HGP Phase 3	9,971	May05/00
HGP Phase 0	3,189	Nov 1/00
HGP Phase 1	20,447	Nov 1/00
HGP Phase 2	1,619	Nov 1/00
HGP Phase 3	9,224	Nov 1/00
HGP Chromosomal assemblies	2759	Aug 1/00

5

**Results:**

The sources for the sequence information used to extend the genes in the provisional patents are listed below. For genes that were extended using Genewise, the accession numbers of the protein homologue and the genomic DNA are given. (Genewise uses the homologue to  
 5 assemble the coding sequence of the target gene from the genomic sequence). The amino acid sequences for the homologs were obtained from the NCBI non-redundant database of proteins (<http://www.ncbi.nlm.nih.gov/Entrez/protein.html>). The genomic DNA came from two sources: Celera and NCBI-NRNA, as indicated below. cDNA sources are also listed below.  
 Abbreviations: HGP: Human Genome Project; NCBI, National Center for Biotechnology  
 10 Information.

SGK187, ID#NA\_1, ENCODING SEQ ID NO:58

Gene name: CRIK

Genewise homologs: AAC72823.1, AAC25483.1,

15 Genomic DNA sources:

Celera 181000000994928; 17000140635257

NCBI AC002563.1, BAC clone 277F10, AC004813

SGK064, ID#NA\_2, ENCODING SEQ ID NO:59

20 Gene name: GRK7 G protein-coupled receptor kinase,

Genewise homolog: AAC9500, gi\_4001826

Genomic DNA sources: Celera, 90000626316497, public hgp contig gi|8139716

Celera 11000284009826, 17000062696662

Incyte cDNA: 7477204CB1 (99% identical)

25 Compared with the homolog, the Celera genewise predicted only the N-terminus; the hgp contig was phase1 and so composed of unordered pieces. It's genewise had two predictions, the second of which extended the prediction to the end of the homolog. A single AA (M) was added to the beginning, based on a Genscan prediction on the Celera contig.

30 SGK409 (SEQ ID NO:3, ENCODING SEQ ID NO:60)

Genewise homologs: NP\_032667, BAA76817

FLv predicted from contig of SGK409 (at 5' end to nucleotide 2193) and KIAA0303 (at 3' end, from 2194 to 77850)

SGK021 (SEQ ID NO:4, ENCODING SEQ ID NO:61)

5 Genewise homologs: CAB76566, BAA95027, CAB76471, NP\_060871

Genomic DNA sources:

Celera 11000283991970, 17000028181153, 17000077758366, 173000019633053, 173000019394610; Incyte cDNA: 1110037.1

Notes: ESTs confirm part of the sequence, but essentially this is a genewise prediction on  
10 90000641092679 using gi\_7161864

SGK410 (SEQ ID NO:5, ENCODING SEQ ID NO:62)

Genewise homolog: NP\_002731 (PKC, iota\_h)

Genomic DNA source: AL133280.12

15 Note: FL from Genewise prediction from HGP contig AL133280.12 using NP\_002731.1

SGK069 (SEQ ID NO:6, ENCODING SEQ ID NO:63)

Genewise homolog: BAA36362

Genomic DNA sources: HGP\_7191033\_7, 11000284155330

20 Celera: 90000641726632

SGK110 (SEQ ID NO:7, ENCODING SEQ ID NO:64)

Genewise homolog: BAA36362

Genomic sources:

25 HGP\_7191033\_7

Celera: 11000283338966, 17000140258105, 17000077607693, 11000283338966 and 17000077607693

SGK053, ID#NA\_8, ENCODING SEQ ID NO:65)

30 Gene name: CKLiK

Genewise homolog: Q63450, NM\_020397.1

Genomic DNA sources: 17000035790284

cDNAs from dbEST: BE266955, AI923704, AW501047

5           SGK124 (SEQ ID NO:9, ENCODING SEQ ID NO:66)

Genewise homolog: gi\_4827159

Genomic DNA source: 90000641832427

Celera:17000047891899,

NCBI: dJ1103G7.3

10   cDNA: Incyte 328225.13

SGK254 (SEQ ID NO:10, ENCODING SEQ ID NO:67)

Genewise homolog: AAB46910

Genomic DNA source: 17000091533743, AC005940

15

SGK297, ID#NA\_11, ENCODING SEQ ID NO:68)

Gene name: CaMKIb2\_h

Genomic DNA source:

Celera 17000140614482\_2, 17000113122540

20   cDNA sources: Incyte 827431CB1; dbEST R87839

SGK411 (SEQ ID NO:12, ENCODING SEQ ID NO:69)

Genewise homolog: AAD20442

Genomic DNA sources: HGP 2828765\_1\_3

25   cDNA sources: NCBI AF071569

SGK027 (SEQ ID NO:13, ENCODING SEQ ID NO:70)

Genewise homologs: AAF69801, AAA97437

FL cloned from adipose and brain

30   Genomic DNA sources:

Celera 17000057443791, 17000029868654, CA2\_GS\_N\_106000011351167\_1  
CA2\_GS\_N\_106000011351167\_1, 17000057443791 on N-term  
cDNA: Incyte 321074.1

SGK027 has been cloned as a full length gene from human brain and adipocyte libraries.

5

SGK046b (SEQ ID NO:14, ENCODING SEQ ID NO:71)

Genewise homologs: AAC33487, AAF69801

Genomic DNA sources: Celera 11000283376057, 11000283376057

10           SGK046c (SEQ ID NO:15, ENCODING SEQ ID NO:72)

Genewise homologs: AAC33487 AAF69801

Genomic DNA sources: Celera 11000284253087, 11000284253087

SGK089 (SEQ ID NO:16, ENCODING SEQ ID NO:73)

15           Genewise homologs: AAF64455, AAA97437

Genomic DNA sources: Celera 11000283986586

cDNA source: NCBI AK024110

SGK133, ID#NA\_17, ENCODING SEQ ID NO:74)

20           Genewise homolog: CAA07196

Genomic DNA sources: 17000075929111, 17000048133019, 17000076096636, 10  
17000140484696

cDNA sources: NCBI N83965; Incyte: 999618, 7477486CB1, 984011.1

25           SGK004 (SEQ ID NO:18, ENCODING SEQ ID NO:75)

Genewise homolog: gi|9978891

Gene name: MSK, SIK (AB020480, salt-inducible protein kinase),

Genomic DNA sources: Celera 17000084574278, 17000048604376, 78000006706941

30           SGK006 (SEQ ID NO:19, ENCODING SEQ ID NO:76)

Genewise homolog: NP\_056570, gi|7657216.

Genomic DNA sources: Celera 11000283913252, 78000005691234

SGK180, ID#NA\_20, ENCODING SEQ ID NO:77)

5 Gene name: SNRK\_h

Genewise homolog: gi|7303211

Genomic DNA sources: Celera 17000057577785, 181000001006215

cDNA sources: NCBI AF226044, AK000231 (N-term) , D43636 KIAA0096 (C-term)

10 SGK386, ID#NA\_21, ENCODING SEQ ID NO:78)

Gene name: MLCKs\_h

Genomic DNA sources:

Celera 17000140249749\_2, 17000140438265

HGP: 7242443\_3, AL160175, AL160175.5

15

SGK003 (SEQ ID NO:22, ENCODING SEQ ID NO:79)

Genewise homolog: P70065, gi|6166236

Genomic DNA sources:

Celera 17000083956390, 17000062613386, 17000036890480, 780000063066170,

20 78000006306170

FL genewise prediction from Celera sequence using CKIalpha P70065 as a model; manually corrected by addition of seq gataactaa at 1006-1014 based on pairwise comparison between aa seq for SGK003 and CKIalpha\_h NP\_001883.2 and CKI, epsilon\_h A57011. Ctermimus of SGK003 confirmed with HGP AL391383.4. Genscan predicts a similar protein from the same  
25 contig, with a stop two AAs after the end of the genewise prediction, so those two AAs (DN) were added to the genewise prediction, which then made it the same length as it's closest public homolog.

SGK066 (SEQ ID NO:23, ENCODING SEQ ID NO:80)

30 Genewise homolog: T42260



Genomic DNA source: 11000283296349

cDNA source: dbEST AA234451

SGK041 (SEQ ID NO:24, ENCODING SEQ ID NO:81)

5 Gene name: NKIAMRE

Genomic DNA sources: Celera 17000062743907, 17000076002106, 17000048152347,  
17000091635260

cDNA sources: AF130372

Genewise was carried out on Celera contig 300702668 using protein homologs gi\_4505569,  
10 gi\_7001374, and gi\_7706059. Genscan predicts an extension to the C-terminus, which extends  
the homology to a rat homolog further. The last 4 AA of the genewise prediction disagree with  
the genscan prediction and were removed. The Genscan prediction is supported by NCBI ESTs:  
gi|10823329, gi|2963750, gi|5111720. All lose homology with the genscan at the same point,  
where we believe the genscan prediction to be incorrect. The C-terminal end of the gene was  
15 predicted using these EST sequences along with Celera genomic sequence, and the open reading  
frame extended down to the first verified stop.

Genewise shows that this gene has 12 exons, as follows (co-ordinates are from the predicted  
cDNA; the kinase domain is 4-286)

1-165, 166-360, 361-539, 540-652, 653-792, 793-881, 882-1035, 1036-1364, 1365-1458, 1459-  
20 1621, 1622-1719, 1720-1773.

Many ESTs align to the gene, but are missing one or more exons, showing a large degree of  
alternative splicing in this gene

3943561T8 and 3943561F8 read through the intron between exons 7 and 8. This is unlikely to be  
functional, as it shifts the frame and encodes stop codons, but might result in a truncated protein.

25 Public EST gi|2849359 is missing exon 3

Incyte EST 6553714H1 is missing exon 5

Incyte EST 6442763H1 is missing exon 6

Incyte EST 918059R1 is missing exon 7

Incyte ESTs 1436805F1, 1436805H1, 1436805F6, 1436805F1 and 1436805H1, and public ESTs  
30 gi|677041 and gi|776279 are missing exons 5 and 6

Public EST gi|1521237 is missing exons 5-7

Incyte EST 6859179H1 and public EST gi|5368476 are missing exons 6 and 7

Public EST gi|4988790 is missing exons 6-9

Public EST gi|2963750 is missing exons 7 & 8

- 5 The public NKIAMRE sequence of this gene, and many ESTs read through the end of exon 8 into intron 8 and encode a shorter protein form. IncyteESTs 2494301F6 and 2494301H1 continue from exon4 into intron 4 before splicing out to exon5, and so introduce some extra sequence into the protein

```

1, bp=n TLAAPGDIYTDYVATRWRAPVLKDTSYGK-----PVDIWALGCMIIEMAT
10      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1, bp=n TLAAPGDIYTDYVATRWRAPVLKDTSYGKYVYFGILSAFPRPVDIWALGCMIMRWPL

```

#### SGK112 (SEQ ID NO:25, ENCODING SEQ ID NO:82)

Genewise homolog: S22745, gi\_107655

- 15 Genomic DNA source: Celera 17000035915087, 300960782

cDNA sources: Incyte 1698381CB1

Comparison with incyte EST 1698381CB1 gives a C-terminal extension, confirmed with further genewise. The contig was assembled and extended by the addition of the following ESTs:

- 1698381CB1, 058298.1, 5314910H1, 1698381T6, 1698381F6, and 2539246. The contig was  
 20 confirmed by comparison to genomic sequences in Celera contigs 173000022366173,  
 173000022366176, 173000022366174, and 300968782.

#### SGK038 (SEQ ID NO:26, ENCODING SEQ ID NO:83)

Gene name: ERK7

- 25 Genewise homolog: AAD12719

Genomic DNA sources: 17000030278391, 17000057882051, 17000084204744,  
 11000283886869,

cDNA sources: Incyte 253132.48 gave start Met; Incyte 2613981F6 gave stop.

- Genewise with 142000016404854 and gi\_4220888. Comparison with NCBI EST gi|9510335  
 30 adds on 3AA at the end to bring it to a stop, and incyte EST adds a short N-terminal extension.

The internal sequence "DMGFLLAPPTHTPVFLSLQ" is not predicted by genewise, and is not seen in another EST (incyte 253132.50) and appears to be an alternative splice form.

SGK158 (SEQ ID NO:27, ENCODING SEQ ID NO:84)

5 Genewise homologs: Q60936 Q92338

CDNA source: Incy406057\_19

Genewise with hgp contig gi|9797056 and homologs gi|3025215, gi|7292815. The result is identical in part to a number of hypothetical proteins from human and mouse. Incyte EST 406057.19 extends and confirms the full length sequence.

10

SGK429 (SEQ ID NO:28, ENCODING SEQ ID NO:85)

Genewise homolog: CAB76039

Genomic DNA sources: HGP\_5001549\_3;

cDNA sources: Incyte 052560.1

15 Notes: Genewise with 90000642611957 and homologs gi|5441947, gi|7106068, gi|7331802.

Prediction is extended and verified by NCBI ESTs: gi|10326317, gi|10159451, gi|8278227, gi|9096664, gi|9096207, gi|570679, gi|10223903.

SGK152 (SEQ ID NO:29, ENCODING SEQ ID NO:86)

20 Gene name: SUDD

Genomic DNA source: 11000258249295

Genewise with 90000641287353 and gi|7295659, extended with NCBI ESTs gi|10952245, gi|1099151 gi|10998012, gi|9892217, gi|9772905.

25

SGK077 (SEQ ID NO:30, ENCODING SEQ ID NO:87)

Genewise homolog: NP\_034483, gi\_7106329

Genomic DNA sources: HGP BAC\_AF168787; Celera 11000283672012, 66000026756418 17000057552303

30 CDNA sources: Incyte 068072.1

Note: When checked against EST database, a short stretch of the genewise prediction was shown not to match several ESTs (incyte template 068072.1, LGfl 387382CB1 and LGflft sequence 387382CB1). The polymorphism occurs in the middle of an exon. The polymorphic variants differ in the following region.

5    SGK077\_1    RPPQKCSTPCGPLRLPPFPSRDGRLSPDLVCGQPRDGDELGISASLFSSSLASPCPGSP 60  
                  RPPQKCSTP            P P+            LSPDLVCGQPRDGDELGISASLFSSSLASPCPGSP  
      SGK077\_2    RPPQKCSTPAARSDFRPSAAIRP-LSPDLVCGQPRDGDELGISASLFSSSLASPCPGSP 901081

SGK093, Wnk3 (SEQ ID NO:31, ENCODING SEQ ID NO:88)

10    Genewise homolog: AAF74258

Genomic DNA sources: Celera 11000284063441, 11000284266180; HGP

CA2\_GS\_N\_111000004311351\_2

Note: FLv from Genscan prediction from Celera 96000001680843; corrected manually by deletion of aa sequence

15    "HRLGECWQKMRRRQQGAAGGNFPVGGSPEDVSPHQDSGYAPSPR" which did not fit the HMM prediction and did not align to NEK7. Genscan prediction supported by multiple, overlapping cDNA sequences (Incyte 8036923J1, cluster 77056\_1 Inc, NCI CGAP EST BF000369.1, Incyte 7934590H1, Incyte 604878F6 and cluster 806401\_1). The translated sequence was compared to the Genscan database of gene predictions from Celera Assembly  
 20    release 3, and was found to match a predicted coding sequence from contig 96000001680843. Comparison of the predicted structure with other members of this family led to a deletion of predicted amino acids 253-297 in the putative catalytic domain. The upstream sequence was confirmed and extended with EST sequences 7459014H1, 2683093F6, 1001757.2, and 6789813F8. Two additional exons with a total of 177 nucleotides were added at amino acid  
 25    position 655 (of the original Genscan prediction) based on ESTs 1001757.2 and 6789813F8. The sequences between amino acid 739 and 770 of the original prediction were replaced by a section of 697 nucleotides, based on ESTs 6622825J1, 715385, 7605034J1, 197338H1, 7609905H1, 7613564H1, 3734294F6, and 5290816T6. The remainder of the sequence was confirmed and extended with by the addition of EST 475276.9, 3038391H1, and 2963959.

30

SGK074 (SEQ ID NO:32, ENCODING SEQ ID NO:89)

Genewise homolog: NP\_034562

Genomic DNA sources:

Celera 11000283476699, 17000113567391;

HGP 6758860\_1\_5

5           SGK087 (SEQ ID NO:33, ENCODING SEQ ID NO:90)

Genewise homologs: AAD47290, NP\_003574, gi\_4503427, gi\_5702386

Genomic DNA sources: Celera 82000011582439

Celera 11000283914919, 17000036171996,

HGP BAC\_AC005832

10

          SGK295, KIS (SEQ ID NO:34, ENCODING SEQ ID NO:91)

Genewise homolog: NP\_058989 (KIS\_r), X98374

Genomic DNA sources: Celera 17000113079883

15           SGK419 (SEQ ID NO:35, ENCODING SEQ ID NO:92)

Genewise homologs: CAB70863, NP\_055726

Genomic DNA sources: HGP\_1 DKFZp434P0116\_h

CDNA sources: NM\_017593

Notes: the contig is 92000005058101 and the genewise template is gi\_6807781.

20

          SGK125, MYO3A (SEQ ID NO:36, ENCODING SEQ ID NO:93)

Gene name: MYO3A

Genomic DNA sources: Celera 17000047912903, 11000502322294, 17000078090910,

25   17000077958351, 39000026222925,

CDNA source: NM\_017433.1

          SGK445 (SEQ ID NO:37, ENCODING SEQ ID NO:94)

Genewise homolog: NP\_055079

30   Genomic DNA source:

Celera: 59000028993040, 4000001803382,

NCBI: 5420628

SGK127 (SEQ ID NO:38, ENCODING SEQ ID NO:95)

5 Genewise homolog: NP\_038599

Genomic DNA sources: Celera 17000084323304, 17000047957940, 17000140021687,  
96000001217832\_1, 17000062804843.

CDNA sources: Incyte 1041923.1, 7474604CB1

10 Genewise with Celera contig and homolog 90000642658207 and homolog gi|7305215, gave a  
partial prediction. A region similar to the unmatched part of the homolog could be genewised  
with the same homolog from 90000642658125. Incyte EST 7474604CB1 confirmed the full  
clone.

SGK009, ANKRD3 (SEQ ID NO:39, ENCODING SEQ ID NO:96)

15 Genewise homolog: gi\_7768736, gi\_7768736

Genomic DNA source: AP001743 (Homo sapiens genomic DNA, chromosome 21q),  
AP001743.1, contig 78000006822431. Splice variant:

1, bp=n RPELPPVCRARPRACSHLIRLMQRCWQGDPRVRPTFQ-----

.....

20 SGK009 RPELPPVCRARPRACSHLIRLMQRCWQGDPRVRPTFQNGNLNGELIRQVLAALLPVTGRW  
250 260 270 280 290 300

280 290 300 310

1, bp=n -----EITSETEDLCEKPDDEVKETAHDLVKSPPEPRSE

25 SGK009 RSPGEGFRLESEVIIRVTCPLSSPQEITSETEDLCEKPDDEVKETAHDLVKSPPEPRSE  
.....

SGK421 (SEQ ID NO:40, ENCODING SEQ ID NO:97)

30 Gene name: STK22A, TSK1

Genewise homolog: NP\_033461

Genomic DNA sources: Celera 4000001803622;

SGK047 (SEQ ID NO:41, ENCODING SEQ ID NO:98)

Genewise homolog: NP\_004320

Genomic DNA source: Celera 11000258262374

5           SGK196 (SEQ ID NO:42, ENCODING SEQ ID NO:99)

Genewise homologs: T01289, AA210451

Genomic DNA source: 17000062537825

cDNA source: AK027009.1

10           SGK396 (SEQ ID NO:43, ENCODING SEQ ID NO:100)

Genewise homologs: CAB90410, CAA11152

Genomic DNA sources:

HGP: 5630059\_1\_4, AL048858, HGP\_1\_5630059\_1\_4

Celera: 17000091618909, 173000013978058

15

SGK279, ID#NA\_44, ENCODING SEQ ID NO:101)

Gene Name: PKN

Genewise homologs: BAA36362, S71887

Genomic DNA sources: Celera 11000508234504, 17000091439256, 17000097259742,

20   17000084154777

CDNA sources: Incyte 7312543CB1

SGK037, ID#NA\_45, ENCODING SEQ ID NO:102)

Genewise homolog: P51956, P51954

25   Genomic DNA sources: 17000036048987, 17000030265658, 17000062680964.

Note: FL virtual from genewise/Genscan of Celera assembly 173000036274838; "MDK" start from HGP 6996171 (positions 144536-114544). Cterm predicted from Genscan of Celera assembly 173000036274838.

30           SGK060 (SEQ ID NO:46, ENCODING SEQ ID NO:103)

Genewise homolog: NP\_035979, gi\_6754822

Genomic DNA sources: Celera 17000036897142, 84000006837210

Genewise gave a short prediction, which was extended with ESTs: NCBI gi|9334037, gi|1690632, gi|1471842, gi|4196791, gi|8149570 and Incyte sequence 7477585CB1. Incyte  
5 sequence 281154.4 shows an alternative splice form, missing

"SLACILYEMCCMNHAFAGSNFLSIVLKIVEGDTPSLPERYPKELNAIMESMLNKNPSLRPSAIEILKIPYLDEQLQNLRCRYSEM  
TLEDKNLDCQKEAAHIINAM" from the sequence at -648-962 on the predicted cDNA.

SGK080 (SEQ ID NO:47, ENCODING SEQ ID NO:104)

10 Genewise homolog: NP\_002488

Genomic DNA sources:

Celera:17000048119006, 11000283866229, 17000084037705, 17000084478382,  
17000084564941, 17000096840584, 17000062653550, 17000112945690,

HGP: BAC\_AP000532 = genomic DNA, chromosome 22q11.2, Cat Eye Syndrome region

15

SGK002 (SEQ ID NO:48, ENCODING SEQ ID NO:105)

Genewise homolog: P36507

Genomic DNA sources: 11000501691092,

Note: FL from Genewise prediction from HGP AC018639.8 using MPK2\_h P36507 as a  
20 template.

SGK058 (SEQ ID NO:49, ENCODING SEQ ID NO:106)

Genewise homologs: A48084, AAC97114

Genomic DNA sources:

25 Celera 17000036225371, 111000004304440

HGP: AK026727.1

cDNA source: Incyte 217301.4

SGK103 (SEQ ID NO:50, ENCODING SEQ ID NO:107)

30 Genewise homolog: CAA39285, AAA28552



Genomic DNA source: Celera 11000284272557

SGK035 (SEQ ID NO:51, ENCODING SEQ ID NO:108)

Genewise homolog: Q13177

5 Genomic DNA sources: Celera 17000030169905, 39000025586693, 39000025680595,  
92000003842663

HGP: gi\_3041712

SGK075 (SEQ ID NO:52, ENCODING SEQ ID NO:109)

10 Genewise homologs: NP\_059129 P50527

Genomic DNA sources: Celera 11000283492249, 11000284453752, 39000025994824,  
17000113802883

Notes: Genewise with gi\_10440888 and 301266624, and 334000009836768 with gi\_10440888,  
along with NCBI EST gi|8362435, followed by Genewise on 301266624 using a composite

15 prediction as template. The incyte EST is missing the bolded sequence. An alternative splice  
was noted with the following sequence inserted after bp 933 sequence:

ccctatgctgaacagagggactacaaaggcaaatcagatgcagttcctgataaagagctaatagtctgg. This insert maintains  
the same reading frame, with the inset coding for the following peptide sequence:

PYAEQRDYKGKSDAVPDKELIVW. This insert occurs after amino acid 311 (i.e., after the R

20 in the following sequence: KHQNPVAKTR[PYAEQRDYKGKSDAVPDKELIVW]).

SGK188, (SEQ ID NO:53, ENCODING SEQ ID NO:110))

Gene name: EphA9

Genomic DNA sources: 17000057739161, 17000065041969, 17000062798377;

25 cDNA sources: Incyte 7474721CB1

SGK040 (SEQ ID NO:54, ENCODING SEQ ID NO:111)

Genewise homologs: P41243, T33475, gi\_729890 and gi\_7508561

Genomic DNA sources: Celera 11000257912897; HGP 11000257912897, contig

30 173000019399041.

Genewise prediction was extended by ESTs and an open reading frame identified. The ESTs were: NCB sequences gi|10722519, gi|1920438, gi|1927810, gi|8656982, gi|4609396, gi|6087575 and Incyte sequences 215217.8, 215217.7, 1242491CB1

5           SGK390 (SEQ ID NO:55, ENCODING SEQ ID NO:112)

Genewise homolog: Q64398

Genomic DNA sources: 17000077925509, 17000077861414

CDNA sources: 7474637CB1

10           SGK007 (SEQ ID NO:56, ENCODING SEQ ID NO:113)

Genewise homolog: T42260

Genomic DNA sources: Celera 17000113227249

            SGK050 (SEQ ID NO:57, ENCODING SEQ ID NO:114)

15           Genewise homolog: P16067

Genomic DNA source: Celera 17000035767558

            SGK187, CRIK (SEQ ID NO:1, ENCODING SEQ ID NO:58) is 6159 nucleotides long. The open reading frame starts at position 1 and ends at position 6159, giving a ORF length of  
 20   6159 nucleotides. The predicted protein is 2053 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, AGC, DMPK. This gene maps to chromosomal position 12q24.23. Amplification of this chromosomal position has been associated with the following human diseases: Non-small cell lung cancer (12q24.1-24.3; 2/50) (Knuutila, et al.). This gene contains candidate single  
 25   nucleotide polymorphisms at the following positions: 2874=R (ss1337340); 2883=Y (rs904655); 3327=R (ss1581624). ESTs for this gene in the public domain (dbEST) are: BE909486, BE875297, AW605350. This gene has repetitive sequence at the following nucleotide positions: none.

SGK064, GRK7 (SEQ ID NO:2, ENCODING SEQ ID NO:59) is 1662 nucleotides long. The open reading frame starts at position 1 and ends at position 1659, giving a ORF length of 1659 nucleotides. The predicted protein is 553 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, AGC, GRK. This gene maps to chromosomal position 3q24. This chromosomal position has been associated with the following human diseases: squamous cell carcinomas of the head and neck (3/30) and Uterine cervix cancer (3/10), (Knuutila, et al.), Usher syndrome (OMIM, 276902 USHER SYNDROME, TYPE III; USH3). This gene contains candidate single nucleotide polymorphisms at the following positions: 965=K; 1318=R. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: 295 to 314.

SGK409, KIAA0303 (SEQ ID NO:3, ENCODING SEQ ID NO:60) is 7785 nucleotides long. The open reading frame starts at position 1 and ends at position 7569, giving a ORF length of 7569 nucleotides. The predicted protein is 2523 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, AGC, MAST. This gene maps to chromosomal position 5q12.1. This chromosomal position has been associated with the following human diseases: cancer of the testis (15q15-qter; 2/11) (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 6282=M ; 6327=M. ESTs for this gene in the public domain (dbEST) are: BE515326, BE267294, AA926642. This gene has repetitive sequence at the following nucleotide positions: 2355 to 2375.

SGK021 (SEQ ID NO:4, ENCODING SEQ ID NO:61) is 981 nucleotides long. The open reading frame starts at position 1 and ends at position 981, giving a ORF length of 981 nucleotides. The predicted protein is 327 amino acids long. This sequence contains the entire catalytic region of a novel kinase. It is classified as (superfamily/group/family): protein kinase, AGC, Mo3C11.1\_ce. This gene maps to chromosomal position 5q31.2. This chromosomal position has been associated with the following human diseases: Chondrosarcoma (2/45), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the

following positions: 9=S; 97=R. ESTs for this gene in the public domain (dbEST) are: none.  
This gene has repetitive sequence at the following nucleotide positions: none.

5       SGK410 (SEQ ID NO:5, ENCODING SEQ ID NO:62) is 2260 nucleotides long. The  
open reading frame starts at position 72 and ends at position 1964, giving a ORF length of 1893  
nucleotides. The predicted protein is 631 amino acids long. This sequence is full length (start  
methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, AGC,  
PKC. This gene maps to chromosomal position Xq23. This chromosomal position has been  
associated with the following human diseases: cancer of the prostate (Xq23-qter; 1/9),  
10   (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the  
following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none.  
This gene has repetitive sequence at the following nucleotide positions: none.

15       SGK069 (SEQ ID NO:6, ENCODING SEQ ID NO:63) is 1716 nucleotides long. The  
open reading frame starts at position 1 and ends at position 1716, giving a ORF length of 1716  
nucleotides. The predicted protein is 572 amino acids long. This sequence contains the entire  
catalytic region of a novel kinase. It is classified as (superfamily/group/family): protein kinase,  
AGC, Unique. This gene maps to chromosomal position 19p11-p13. This chromosomal  
position has been associated with the following human diseases: Small cell lung cancer (19p12,  
20   2/22) (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the  
following positions: 1180=S (ss1317629); 210=Y (ss1688813). ESTs for this gene in the public  
domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide  
positions: 1656 -1678.

25       SGK110 (SEQ ID NO:7, ENCODING SEQ ID NO:64) is 1119 nucleotides long. The  
open reading frame starts at position 1 and ends at position 1119, giving a ORF length of 1119  
nucleotides. The predicted protein is 373 amino acids long. This sequence is a partial kinase  
catalytic domain. It is classified as (superfamily/group/family): protein kinase, AGC, Unique.  
This gene maps to chromosomal position 19q13.4. This chromosomal position has been  
30   associated with the following human diseases: cancer of the breast (19q13.1-qter; 1/33),

(Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 597=R (rs654439); 252=Y (ss661406). ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

5

SGK053, CKLiK (SEQ ID NO:8, ENCODING SEQ ID NO:65) is 1074 nucleotides long. The open reading frame starts at position 1 and ends at position 1071, giving a ORF length of 1071 nucleotides. The predicted protein is 357 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CAMK, AMPK. This gene maps to chromosomal position 10p14. This chromosomal position has been associated with arrhythmogenic right ventricular dysplasia (OMIM, 604401 ARRHYTHMOGENIC RIGHT VENTRICULAR DYSPLASIA, FAMILIAL). This gene contains candidate single nucleotide polymorphisms at the following positions: 605=Y; 509=M. ESTs for this gene in the public domain (dbEST) are: BE266955, AI923704, AW501047. This gene has repetitive sequence at the following nucleotide positions: 419 to 440.

15

SGK124 (SEQ ID NO:9, ENCODING SEQ ID NO:66) is 1077 nucleotides long. The open reading frame starts at position 1 and ends at position 1074, giving a ORF length of 1074 nucleotides. The predicted protein is 358 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CAMK, AMPK. This gene maps to chromosomal position 20p12.2-p13. This chromosomal position has been associated with the following human diseases: cancer of the gastroesophageal junction (20p12; 3/28), (Knuutila, et al.); familial noncompaction of left ventricle (OMIM, 604169 NONCOMPACTION OF LEFT VENTRICULAR MYOCARDIUM, FAMILIAL ISOLATED, AUTOSOMAL DOMINANT, 20p13). This gene contains candidate single nucleotide polymorphisms at the following positions: 188=S; 333=Y. ESTs for this gene in the public domain (dbEST) are: BF026145, BE298893. This gene has repetitive sequence at the following nucleotide positions: none.

20

25

SGK254, CAMKka (SEQ ID NO:10, ENCODING SEQ ID NO:67) is 1542 nucleotides long. The open reading frame starts at position 1 and ends at position 1539, giving a ORF length

30

of 1539 nucleotides. The predicted protein is 513 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CAMK, CAMK. This gene maps to chromosomal position 17p13.3. This chromosomal position has been associated with the following human diseases: Lost in cervical cancer (loss of heterogeneity, Lazo, The molecular genetics of cervical carcinoma. Br J Cancer. 1999 Aug;80(12):2008-18. Review). This gene contains candidate single nucleotide polymorphisms at the following positions: 555=R (ss84265); 1148=R. ESTs for this gene in the public domain (dbEST) are: BE783149. This gene has repetitive sequence at the following nucleotide positions: none.

10

SGK297, CaMKIb2 (SEQ ID NO:11, ENCODING SEQ ID NO:68) is 1032 nucleotides long. The open reading frame starts at position 1 and ends at position 1029, giving a ORF length of 1029 nucleotides. The predicted protein is 343 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CAMK, CAMK. This gene maps to chromosomal position Xq28. Translocations involving this chromosomal position has been associated with the following human diseases: human T cell prolymphocytic leukemia (Laine, et al. Mol Cell. 2000 Aug;6(2):395-407); mental retardation (e.g., Russo, et al., Am J Med Genet. 2000 Oct 23;94(5):376-82). Also in Mantle cell lymphoma (Xq26-q28, 5/50). (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 77=K. ESTs for this gene in the public domain (dbEST) are: AI696123, AI141657. This gene has repetitive sequence at the following nucleotide positions: none.

20

SGK411, CaMKII delta2 (SEQ ID NO:12, ENCODING SEQ ID NO:69) is 1500 nucleotides long. The open reading frame starts at position 1 and ends at position 1497, giving a ORF length of 1497 nucleotides. The predicted protein is 499 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CAMK, CAMK. This gene maps to chromosomal position 4q25. This chromosomal position has been associated with the following human diseases: developmental glaucoma (Rieger syndrome, iris hypoplasia, and iridogoniodysgenesis; Craig, et al. Curr Opin

30

Ophthalmol. 1999 Apr;10(2):126-34) This gene contains candidate single nucleotide polymorphisms at the following positions: 15=M; 1387=S (ss1531091). ESTs for this gene in the public domain (dbEST) are: AW502248, AW504981, AA316038. This gene has repetitive sequence at the following nucleotide positions: none.

5

SGK027 (SEQ ID NO:13, ENCODING SEQ ID NO:70) is 1311 nucleotides long. The open reading frame starts at position 1 and ends at position 1308, giving a ORF length of 1308 nucleotides. The predicted protein is 436 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. This gene maps to chromosomal position 5q11-q11.1. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocytoma of bone (1/26) (Knuutila, et al.). Also with B-cell non-Hodgkin's lymphoma. (Wlodarska, et al. Cytogenet Cell Genet. 1994;65(3):179-83). This gene contains candidate single nucleotide polymorphisms at the following positions: 45=Y. ESTs for this gene in the public domain (dbEST) are: BE551926, AI609751, AW138653. This gene has repetitive sequence at the following nucleotide positions: none.

SGK046b (SEQ ID NO:14, ENCODING SEQ ID NO:71) is 225 nucleotides long. The open reading frame starts at position 1 and ends at position 225, giving a ORF length of 225 nucleotides. The predicted protein is 75 amino acids long. This sequence is a partial kinase catalytic domain. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. This gene maps to chromosomal position 3p24.1. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocytoma of soft tissue (3p24-p26, 2/30) (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

SGK046c (SEQ ID NO:15, ENCODING SEQ ID NO:72) is 117 nucleotides long. The open reading frame starts at position 1 and ends at position 117, giving a ORF length of 117

nucleotides. The predicted protein is 39 amino acids long. This sequence is a partial kinase catalytic domain. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. This gene maps to chromosomal position 3p24.1. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocytoma of soft tissue  
 5 (3p24-p26, 2/30) (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

10 SGK089 (SEQ ID NO:16, ENCODING SEQ ID NO:73) is 252 nucleotides long. The open reading frame starts at position 1 and ends at position 252, giving a ORF length of 252 nucleotides. The predicted protein is 84 amino acids long. This sequence is a partial kinase catalytic domain. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. This gene maps to chromosomal position 3p25.3. This chromosomal position has not been  
 15 associated with human diseases. This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

20 SGK133 (SEQ ID NO:17, ENCODING SEQ ID NO:74) is 2385 nucleotides long. The open reading frame starts at position 1 and ends at position 2382, giving a ORF length of 2382 nucleotides. The predicted protein is 794 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. This gene maps to chromosomal position 7p11.2-p21. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocytoma of bone  
 25 (1/26), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 2003=S;1673=S. ESTs for this gene in the public domain (dbEST) are: BE222941. This gene has repetitive sequence at the following nucleotide positions: none.

SGK004, MSK (SEQ ID NO:18, ENCODING SEQ ID NO:75) is 2361 nucleotides long.  
 30 The open reading frame starts at position 1 and ends at position 2358, giving a ORF length of



2358 nucleotides. The predicted protein is 786 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CAMK, EMK-1. This gene maps to chromosomal position 21q22.3. A gene which causes severe ocular alterations and occipital encephalocele (Knobloch syndrome) is mapped to 21q223  
5 (Sertie, et al, Hum Mol Genet 1996 Jun;5(6):843-7). This gene contains candidate single nucleotide polymorphisms at the following positions: 1853=Y (ss571239). ESTs for this gene in the public domain (dbEST) are: AW503500. This gene has repetitive sequence at the following nucleotide positions: none.

10 SGK006 (SEQ ID NO:19, ENCODING SEQ ID NO:76) is 789 nucleotides long. The open reading frame starts at position 1 and ends at position 789, giving a ORF length of 789 nucleotides. The predicted protein is 263 amino acids long. This sequence contains the entire catalytic region of a novel kinase. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK-1. This gene maps to chromosomal position 16q16.1. This chromosomal position  
15 has not been associated with human diseases. This gene contains candidate single nucleotide polymorphisms at the following positions: 4=S (ss1609852). ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

20 SGK180, SNRK (SEQ ID NO:20, ENCODING SEQ ID NO:77) is 2298 nucleotides long. The open reading frame starts at position 1 and ends at position 2295, giving a ORF length of 2295 nucleotides. The predicted protein is 765 amino acids long. This sequence is full length (start methionine to stop codon).It is classified as (superfamily/group/family): protein kinase, CAMK, EMK-1. This gene maps to chromosomal position 3p21.31. This chromosomal position  
25 has been associated with the following human diseases: cancer of the bladder (1/14), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 1817=S. ESTs for this gene in the public domain (dbEST) are: AA447812, AI379954, AA199639. This gene has repetitive sequence at the following nucleotide positions: 1332 to 1353.

30

SGK386, MLCKs (SEQ ID NO:21, ENCODING SEQ ID NO:78) is 1836 nucleotides long. The open reading frame starts at position 1 and ends at position 1836, giving a ORF length of 1836 nucleotides. The predicted protein is 612 amino acids long. This sequence is full length (start methionine to stop codon).It is classified as (superfamily/group/family): protein kinase, CAMK, MLCK. This gene maps to chromosomal position 20q11.1. This chromosomal position has been associated with the following human diseases: papillary renal cell carcinoma. (Lab Invest. 1999 Mar;79(3):311-6). This gene contains candidate single nucleotide polymorphisms at the following positions: 835=M. ESTs for this gene in the public domain (dbEST) are: AA197072, R02824 . This gene has repetitive sequence at the following nucleotide positions:

10 none.

SGK003 (SEQ ID NO:22, ENCODING SEQ ID NO:79) is 1014 nucleotides long. The open reading frame starts at position 1 and ends at position 1014, giving a ORF length of 1014 nucleotides. The predicted protein is 337 amino acids long. This sequence is full length (start methionine to stop codon).It is classified as (superfamily/group/family): protein kinase, CKI, CKI. This gene maps to chromosomal position 13q14.11. This chromosomal position has been associated with the following human diseases: fallopian tube cancer (13q14-qter; 1/12) (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none.

20 This gene has repetitive sequence at the following nucleotide positions: none.

SGK066 (SEQ ID NO:23, ENCODING SEQ ID NO:80) is 1200 nucleotides long. The open reading frame starts at position 1 and ends at position 1200, giving a ORF length of 1200 nucleotides. The predicted protein is 400 amino acids long. This sequence contains the entire catalytic region of a novel kinase. It is classified as (superfamily/group/family): protein kinase, CKI, CKI. This gene maps to chromosomal position 15q15. This chromosomal position has been associated with the following human diseases: cancer of the testis (2/11), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: AA234451. This gene has

25

30 repetitive sequence at the following nucleotide positions: 486 to 504.

SGK041, NKIAMRE (SEQ ID NO:24, ENCODING SEQ ID NO:81) is 1773 nucleotides long. The open reading frame starts at position 1 and ends at position 1773, giving a ORF length of 1773 nucleotides. The predicted protein is 591 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CMGC, CDK. This gene maps to chromosomal position 5q31.1. This chromosomal position has been associated with the following human diseases: cancer of the digestive tract (5q31-qter), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 1033=R;1284=R; 1181=Y. ESTs for this gene in the public domain (dbEST) are: AI684625, AI694352, AI912547. This gene has repetitive sequence at the following nucleotide positions: none.

SGK112 (SEQ ID NO:25, ENCODING SEQ ID NO:82) is 1083 nucleotides long. The open reading frame starts at position 1 and ends at position 1080, giving a ORF length of 1080 nucleotides. The predicted protein is 360 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CMGC, CDK. This gene maps to chromosomal position CHR2. This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: AA626859. This gene has repetitive sequence at the following nucleotide positions: none.

SGK038, ERK7 (SEQ ID NO:26, ENCODING SEQ ID NO:83) is 1677 nucleotides long. The open reading frame starts at position 1 and ends at position 1674, giving a ORF length of 1674 nucleotides. The predicted protein is 558 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, CMGC, MAPK. This gene maps to chromosomal position na. This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: BE464560, AI049667. This gene has repetitive sequence at the following nucleotide positions: 508 to 528.

SGK158 (SEQ ID NO:27, ENCODING SEQ ID NO:84) is 2256 nucleotides long. The open reading frame starts at position 1 and ends at position 2253, giving a ORF length of 2253 nucleotides. The predicted protein is 751 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Microbial PK, ABC1. This gene maps to chromosomal position 1q42.11-42.2. This chromosomal position has not been associated with human diseases. This gene contains candidate single nucleotide polymorphisms at the following positions: 1752=Y (ss1529336). ESTs for this gene in the public domain (dbEST) are: BE797060, AW006971, AI819411. This gene has repetitive sequence at the following nucleotide positions: 1246 to 1265.

10

SGK429 (SEQ ID NO:28, ENCODING SEQ ID NO:85) is 1881 nucleotides long. The open reading frame starts at position 1 and ends at position 1878, giving a ORF length of 1878 nucleotides. The predicted protein is 626 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Microbial PK, ABC1. This gene maps to chromosomal position 7q34-35. This chromosomal position has been associated with the following human diseases: deafness (Mustapha, et al. Eur J Hum Genet. 1998 May-Jun;6(3):245-50). This gene contains candidate single nucleotide polymorphisms at the following positions: 196=R; 919=Y (ss1549835); 1865=Y (ss1517749). ESTs for this gene in the public domain (dbEST) are: BE877541, BE745459, BE259124. This gene has repetitive sequence at the following nucleotide positions: 1630 to 1651.

15  
20

SGK152, SUDD (SEQ ID NO:29, ENCODING SEQ ID NO:86) is 1560 nucleotides long. The open reading frame starts at position 1 and ends at position 1557, giving a ORF length of 1557 nucleotides. The predicted protein is 519 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Microbial PK, RI01. This gene maps to chromosomal position 18p11.1. This chromosomal position has not been associated with human diseases. This gene contains candidate single nucleotide polymorphisms at the following positions: 972=M. ESTs for this gene in the public domain (dbEST) are: BE621277, AA452706. This gene has repetitive sequence at the following nucleotide positions: none.

25  
30

SGK077 (SEQ ID NO:30, ENCODING SEQ ID NO:87) is 2397 nucleotides long. The open reading frame starts at position 1 and ends at position 2394, giving a ORF length of 2394 nucleotides. The predicted protein is 798 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other, C26C2\_ce. This gene maps to chromosomal position 17p13.3. This chromosomal position has been associated with the following human diseases: Lost in cervical cancer (loss of heterogeneity, Lazo, The molecular genetics of cervical carcinoma. Br J Cancer. 1999 Aug;80(12):2008-18. Review). This gene contains candidate single nucleotide polymorphisms at the following positions: 390=Y (ss1658885); 611=R (ss1629760); 985=Y (ss1629759). ESTs for this gene in the public domain (dbEST) are: AA504563, AW752337. This gene has repetitive sequence at the following nucleotide positions: none.

SGK093, Wnk3 (SEQ ID NO:31, ENCODING SEQ ID NO:88) is 4542 nucleotides long. The open reading frame starts at position 1 and ends at position 4539, giving a ORF length of 4539 nucleotides. The predicted protein is 1513 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other, C26C2\_ce. This gene maps to chromosomal position 17q21.1-2. This chromosomal position has been associated with the following human diseases: cancer of the ovary (17q21-qter, 3/47), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 3279 = K; 4078=M. ESTs for this gene in the public domain (dbEST) are: BF000369, AI650786, AA872586. This gene has repetitive sequence at the following nucleotide positions: none.

SGK074 (SEQ ID NO:32, ENCODING SEQ ID NO:89) is 1065 nucleotides long. The open reading frame starts at position 1 and ends at position 1065, giving a ORF length of 1065 nucleotides. The predicted protein is 355 amino acids long. This sequence contains the entire catalytic region of a novel kinase. It is classified as (superfamily/group/family): protein kinase, Other, DYRK. This gene maps to chromosomal position 19p12-19q13. This chromosomal position has been associated with the following human diseases: Small cell lung cancer (2/22),

(Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

5           SGK087 (SEQ ID NO:33, ENCODING SEQ ID NO:90) is 1887 nucleotides long. The open reading frame starts at position 1 and ends at position 1884, giving a ORF length of 1884 nucleotides. The predicted protein is 628 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other, DYRK. This gene maps to chromosomal position 12p13.3. This chromosomal position has been  
10 associated with the following human diseases: hypertension (Disse-Nicodeme, et al. Am J Hum Genet. 2000 Aug;67(2):302-10). This gene contains candidate single nucleotide polymorphisms at the following positions: 269=R (ss88136). ESTs for this gene in the public domain (dbEST) are: BE243995. This gene has repetitive sequence at the following nucleotide positions: 1264 to 1283.

15           SGK295, KIS (SEQ ID NO:34, ENCODING SEQ ID NO:91) is 1260 nucleotides long. The open reading frame starts at position 1 and ends at position 1257, giving a ORF length of 1257 nucleotides. The predicted protein is 419 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase,  
20 Other, DYRK. This gene maps to chromosomal position 1q23.3. This chromosomal position has been associated with the following human diseases: Hematologic neoplasms (11q23-qter; 1/1), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 355=W. ESTs for this gene in the public domain (dbEST) are: BE895119, AI221234, BE145607. This gene has repetitive sequence at the following nucleotide positions:  
25 none.

          SGK419 (SEQ ID NO:35, ENCODING SEQ ID NO:92) is 1986 nucleotides long. The open reading frame starts at position 1 and ends at position 1983, giving a ORF length of 1983 nucleotides. The predicted protein is 661 amino acids long. This sequence is full length (start  
30 methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other,

NAK. This gene maps to chromosomal position 4q24. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocytoma of soft tissue (14q24-q31; 1/58), (Knuutila, et al.). ESTs for this gene in the public domain (dbEST) are: AI735391, AW967339, AA703517. This gene has repetitive sequence at the following  
5 nucleotide positions: none.

SGK125, MYO3A (SEQ ID NO:36, ENCODING SEQ ID NO:93) is 4848 nucleotides long. The open reading frame starts at position 1 and ends at position 4845, giving a ORF length of 4845 nucleotides. The predicted protein is 1615 amino acids long. This sequence is full  
10 length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other, NinaC. This gene maps to chromosomal position 10p12.32. This chromosomal position has been associated with the following human diseases: Mantle cell lymphoma (10p12-p13; 2/45), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 3145=Y; 3204=Y. ESTs for this gene in the public domain (dbEST) are:  
15 AW196373, AA476697. This gene has repetitive sequence at the following nucleotide positions: none.

SGK445 (SEQ ID NO:37, ENCODING SEQ ID NO:94) is 204 nucleotides long. The open reading frame starts at position 1 and ends at position 204, giving a ORF length of 204  
20 nucleotides. The predicted protein is 68 amino acids long. This sequence is a partial kinase catalytic domain. It is classified as (superfamily/group/family): protein kinase, Other, PLK. This gene maps to chromosomal position na. (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following  
25 nucleotide positions: 183 to 200.

SGK127 (SEQ ID NO:38, ENCODING SEQ ID NO:95) is 2838 nucleotides long. The open reading frame starts at position 1 and ends at position 2835, giving a ORF length of 2835 nucleotides. The predicted protein is 945 amino acids long. This sequence is full length (start  
30 methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other,

RAF. This gene maps to chromosomal position 12q24.21. This chromosomal position has been associated with the following human diseases: cancer of the respiratory tract and of the female genital organs (12q24.2). (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 501=S (ss2005786). ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

SGK009, ANKRD3 (SEQ ID NO:39, ENCODING SEQ ID NO:96) is 2499 nucleotides long. The open reading frame starts at position 1 and ends at position 2496, giving a ORF length of 2496 nucleotides. The predicted protein is 832 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other, RIP. This gene maps to chromosomal position 21q22.3. A gene which causes severe ocular alterations and occipital encephalocele (Knobloch syndrome) is mapped to 21q223 (Sertie, et al, Hum Mol Genet 1996 Jun;5(6):843-7) (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: 427- 452.

SGK421, STK22A, TSK1 (SEQ ID NO:40, ENCODING SEQ ID NO:97) is 1104 nucleotides long. The open reading frame starts at position 1 and ends at position 1101, giving a ORF length of 1101 nucleotides. The predicted protein is 367 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other, STK22A. This gene maps to chromosomal position 5q31.1. This chromosomal position has been associated with the following human diseases: Chondrosarcoma (5q31-q32; 2/45), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 510=M (ss2055126); 279=R (ss2055125). ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.



SGK047 (SEQ ID NO:41, ENCODING SEQ ID NO:98) is 93 nucleotides long. The open reading frame starts at position 1 and ends at position 93, giving a ORF length of 93 nucleotides. The predicted protein is 31 amino acids long. This sequence is a partial kinase catalytic domain. It is classified as (superfamily/group/family): protein kinase, Other, STKR.

5 This gene maps to chromosomal position 10p11.21. This chromosomal position has been associated with the following human diseases: Squamous cell carcinomas of the head and neck (10p11-p13; 1/30), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide

10 positions: none.

SGK196 (SEQ ID NO:42, ENCODING SEQ ID NO:99) is 1053 nucleotides long. The open reading frame starts at position 1 and ends at position 1050, giving a ORF length of 1050 nucleotides. The predicted protein is 350 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other,

15 Unique. This gene maps to chromosomal position na. This gene contains candidate single nucleotide polymorphisms at the following positions: 99=R. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

20

SGK396 (SEQ ID NO:43, ENCODING SEQ ID NO:100) is 1419 nucleotides long. The open reading frame starts at position 1 and ends at position 1416, giving a ORF length of 1416 nucleotides. The predicted protein is 472 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other,

25 Unique. This gene maps to chromosomal position na. This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: AL048857, AL048858, T09068. This gene has repetitive sequence at the following nucleotide positions: none.

SGK279, PKN (SEQ ID NO:44, ENCODING SEQ ID NO:101) is 1275 nucleotides long. The open reading frame starts at position 1 and ends at position 1272, giving a ORF length of 1272 nucleotides. The predicted protein is 424 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, Other, YWY3\_ce. This gene maps to chromosomal position 16q22.3. This chromosomal position has been associated with the following human diseases: Diffuse large cell lymphoma of stomach (16q22-ter; 1/7), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 665=Y. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions:

5

10 none.

SGK037 (SEQ ID NO:45, ENCODING SEQ ID NO:102) is 1947 nucleotides long. The open reading frame starts at position 1 and ends at position 1947, giving a ORF length of 1947 nucleotides. The predicted protein is 649 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, STE, NEK. This gene maps to chromosomal position 13q14.12. This chromosomal position has been associated with the following human diseases: fallopian tube cancer (13q14-qter; 1/12) (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are:

15

20 AA393108. This gene has repetitive sequence at the following nucleotide positions: 886 to 912.

SGK060 (SEQ ID NO:46, ENCODING SEQ ID NO:103) is 1938 nucleotides long. The open reading frame starts at position 1 and ends at position 1935, giving a ORF length of 1935 nucleotides. The predicted protein is 645 amino acids long. This sequence is full length (start methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, STE, NEK. This gene maps to chromosomal position 3q22.1. This chromosomal position has been associated with the following human diseases: Mantle cell lymphoma (13q22-q32; 3/72), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: 1463=W. ESTs for this gene in the public domain (dbEST) are: BE388672,

25

AA412114. This gene has repetitive sequence at the following nucleotide positions: 465 to 488; 1468 to 1487.

SGK080 (SEQ ID NO:47, ENCODING SEQ ID NO:104) is 1338 nucleotides long. The  
5 open reading frame starts at position 1 and ends at position 1338, giving a ORF length of 1338  
nucleotides. The predicted protein is 446 amino acids long. This sequence is full length (start  
methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, STE,  
NEK. This gene maps to chromosomal position 22q11.2. This chromosomal position has been  
associated with the following human diseases: Non-small cell lung cancer (1/50), (Knuutila, et  
10 al.). This gene contains candidate single nucleotide polymorphisms at the following positions:  
1159=R (ss1367671); 422=R (ss1855009). ESTs for this gene in the public domain (dbEST) are:  
none. This gene has repetitive sequence at the following nucleotide positions: none.

SGK002 (SEQ ID NO:48, ENCODING SEQ ID NO:105) is 1829 nucleotides long. The  
15 open reading frame starts at position 327 and ends at position 1523, giving a ORF length of 1197  
nucleotides. The predicted protein is 399 amino acids long. This sequence is full length (start  
methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, STE,  
STE11. This gene maps to chromosomal position 7q32.2. This chromosomal position has been  
associated with the following human diseases: Non-small cell lung cancer 7q32-q35; 1/50),  
20 (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the  
following positions: 1165=R; 983=M. ESTs for this gene in the public domain (dbEST) are:  
none. This gene has repetitive sequence at the following nucleotide positions: none.

SGK058 (SEQ ID NO:49, ENCODING SEQ ID NO:106) is 834 nucleotides long. The  
25 open reading frame starts at position 1 and ends at position 834, giving a ORF length of 834  
nucleotides. The predicted protein is 278 amino acids long. This sequence contains the entire  
catalytic region of a novel kinase. It is classified as (superfamily/group/family): protein kinase,  
STE, STE11. This gene maps to chromosomal position 2q21.2. This chromosomal position has  
been associated with the following human diseases: bladder carcinoma (12q21-q24; 1/16),  
30 (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the

following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: 733 to 751.

5       SGK103 (SEQ ID NO:50, ENCODING SEQ ID NO:107) is 84 nucleotides long. The open reading frame starts at position 1 and ends at position 84, giving a ORF length of 84 nucleotides. The predicted protein is 28 amino acids long. This sequence is a partial kinase catalytic domain. It is classified as (superfamily/group/family): protein kinase, STE, STE11. This gene maps to chromosomal position 5p14.3. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocytoma of soft tissue  
10 (6/88), (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

15       SGK035 (SEQ ID NO:51, ENCODING SEQ ID NO:108) is 1044 nucleotides long. The open reading frame starts at position 1 and ends at position 1044, giving a ORF length of 1044 nucleotides. The predicted protein is 348 amino acids long. This sequence is a partial kinase catalytic domain. It is classified as (superfamily/group/family): protein kinase, STE, STE20. This gene maps to chromosomal position CHR15. This gene contains candidate single nucleotide polymorphisms at the following positions: 2273=Y. ESTs for this gene in the public  
20 domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: 133 to 151; 695 to 713.

25       SGK075 (SEQ ID NO:52, ENCODING SEQ ID NO:109) is 3318 nucleotides long. The open reading frame starts at position 1 and ends at position 3318, giving a ORF length of 3318 nucleotides. The predicted protein is 1106 amino acids long. This sequence contains the entire catalytic region of a novel kinase. It is classified as (superfamily/group/family): protein kinase, STE, STE20. This gene maps to chromosomal position 2q31.1. This chromosomal position has been associated with the following human diseases: Squamous cell carcinomas of the head and neck (2q31-q33; 3/30), (Knuutila, et al.). This gene contains candidate single nucleotide  
30 polymorphisms at the following positions: 889=R. ESTs for this gene in the public domain

(dbEST) are: none. This gene has repetitive sequence at the following nucleotide positions: none.

SGK188, EphA9 (SEQ ID NO:53, ENCODING SEQ ID NO:110) is 3112 nucleotides  
5 long. The open reading frame starts at position 74 and ends at position 3100, giving a ORF  
length of 3027 nucleotides. The predicted protein is 1009 amino acids long. This sequence is  
full length (start methionine to stop codon). It is classified as (superfamily/group/family):  
protein kinase, TK, RTK-11. This gene maps to chromosomal position 1p34.1-34.3. This  
chromosomal position has been associated with the following human diseases: cancer of the  
10 testis (1p34-pter; 1/11), (Knuutila, et al.). This gene contains candidate single nucleotide  
polymorphisms at the following positions: 2104=Y (ss1986120); 2319=R. ESTs for this gene in  
the public domain (dbEST) are: none. This gene has repetitive sequence at the following  
nucleotide positions: none.

15 SGK040 (SEQ ID NO:54, ENCODING SEQ ID NO:111) is 2730 nucleotides long. The  
open reading frame starts at position 1 and ends at position 2727, giving a ORF length of 2727  
nucleotides. The predicted protein is 909 amino acids long. This sequence is full length (start  
methionine to stop codon). It is classified as (superfamily/group/family): protein kinase, TK,  
Unique. This gene maps to chromosomal position 12q12. This chromosomal position has been  
20 associated with the following human diseases: Diffuse large cell lymphoma (2/66), (Knuutila, et  
al.). This gene contains candidate single nucleotide polymorphisms at the following positions:  
1869 = R, 1004 = Y. ESTs for this gene in the public domain (dbEST) are: BE177830,  
AF114068, AA283608. This gene has repetitive sequence at the following nucleotide positions:  
460 to 479.

25

SGK390 (SEQ ID NO:55, ENCODING SEQ ID NO:112) is 3495 nucleotides long. The  
open reading frame starts at position 1 and ends at position 3492, giving a ORF length of 3492  
nucleotides. The predicted protein is 1164 amino acids long. This sequence is full length (start  
methionine to stop codon). It is classified as (superfamily/group/family): protein kinase -like,  
30 DAG kin, DAG kin. This gene maps to chromosomal position 13q14.2. This chromosomal

position has been associated with the following human diseases: fallopian tube cancer (13q14-  
 qter; 1/12) (Knuutila, et al.). This gene contains candidate single nucleotide polymorphisms at  
 the following positions: 1314=R. ESTs for this gene in the public domain (dbEST) are:  
 BE715967, AI333565, AW052032. This gene has repetitive sequence at the following  
 5 nucleotide positions: none.

SGK007 (SEQ ID NO:56, ENCODING SEQ ID NO:113) is 2652 nucleotides long. The  
 open reading frame starts at position 1 and ends at position 2652, giving a ORF length of 2652  
 nucleotides. The predicted protein is 884 amino acids long. This sequence contains the entire  
 10 catalytic region of a novel kinase. It is classified as (superfamily/group/family): protein kinase -  
 like, GCyc, GCyc. This gene maps to chromosomal position 10q26.11. This chromosomal  
 position has not been associated with human diseases. (Knuutila, et al.). This gene contains  
 candidate single nucleotide polymorphisms at the following positions: none detected. ESTs for  
 this gene in the public domain (dbEST) are: none. This gene has repetitive sequence at the  
 15 following nucleotide positions: none.

SGK050 (SEQ ID NO:57, ENCODING SEQ ID NO:114) is 144 nucleotides long. The  
 open reading frame starts at position 1 and ends at position 144, giving a ORF length of 144  
 nucleotides. The predicted protein is 48 amino acids long. This sequence is a partial kinase  
 20 catalytic domain. It is classified as (superfamily/group/family): protein kinase -like, GCyc,  
 GCyc. This gene maps to chromosomal position 9p13.1-p13.2. This chromosomal position has  
 been associated with the following human diseases: Non-small cell lung cancer (1/50) and  
 testicular cancer (4/11) (Knuutila, et al.). This gene contains candidate single nucleotide  
 polymorphisms at the following positions: none detected. ESTs for this gene in the public  
 25 domain (dbEST) are: none. This gene has repetitive sequence at the following nucleotide  
 positions: none.

#### **EXAMPLE 2a: Probe Generation**

Genomic fragments were PCR cloned to be used as probes. Exon fragments were cloned  
 30 from genomic DNA sources (HUVEC or HMEC) by PCR methodology. Annealing temperature

for the oligos used was 68C in a 50 microliter reaction. 5 microliter aliquots were analyzed by agarose gel electrophoresis to verify the correct size fragment was obtained. Fragments of the correct size were excised from the agarose gel and gene-cleaned and subcloned into the pCR-TOPO4 vector (Stratagene). These ligated plasmids were transformed into E.coli bacteria (TOP10 strain/ Stratagene) and selected using ampicillin antibiotic resistance. Resulting colonies, four per construct, were grown in media containing ampicillin and the plasmid DNA purified. Restriction digest analysis was carried out to verify the correct DNA insert. Plasmids containing fragments of the correct size were sequence verified using the T7 and T3 primers. DNA fragments to be used as probes for the blots were generated by restriction digest and purification of these fragments. These fragments were radioactively labeled by incorporating alpha-32P dCTP using a Boehringer Mannheim Random Hexamer Labeling Kit. Incorporation of radioactive isotope resulted in probes generated with between  $5 \times 10^7$  to  $2 \times 10^8$  cpm per 100 nano-grams of DNA fragment. The PCR primers used to clone the fragments are listed below:

SGK003.5' GACGTTTATCTGGGCATCACCACC  
 15 SGK003.3' TCGGACAGTGCCAATGAGGTGTTT  
 SGK074.5' GTGATTGACTTCGGATCCGCCAGC  
 SGK074.3' CTTTTCCTTACCTTCGTCTCGGCCAG  
 SGK077.5' CCGGGACCTGGGAGCCGGCTTTTC  
 SGK077.3' GCTCGGGACGGCTGAGGCTGCAAC  
 20 SGK124.5' AAGAAGCTGGTGCTGGAGAAGCTG  
 SGK124.3' AACCACTTCTCTGTCTCCCTCCTC  
 SGK187.5' TCCGACACCATAACTGAGTTACAG  
 SGK187.3' CTGCTCCTGGGCCAATAAAGC  
 SGK386.5' CCTGATGGGTGTCTCACCTCCTCT  
 25 SGK386.3' GGAGATGATGGCAGGACAGCTGGG  
 SGK396.5' TTGGAGGGCTCAGAAGAGGAC  
 SGK396.3' AACAAGTCCCTCATCTCCAGGTGA

30 **EXAMPLE 2b: Expression Analysis of Polypeptides of the Invention**

The gene expression patterns for selected genes were studied using four techniques: 1) a tissue microarray developed at Sugan, containing over 450 tissues and probed with labeled genes; 2) a PCR screen of 48 human tissues (this technique does not yield quantitative expression levels between tissues, but does identify those tissues that express the gene at a level detectable by PCR, as well as those that do not express the gene at such a level), 3) a commercial array of tissue from Clontech, probed with labeled genes, and 4) for one gene (SGK093), an analysis form Northern blotting.

#### 1) Tissue Arrays

"cDNA libraries" derived from over 450 tissue or cell line sources were immobilized onto nylon membranes and probed with <sup>32</sup>P-labeled cDNA fragments derived from the gene(s) of interest. To make the cDNA, total RNA or mRNA was used as template in a reverse transcription reaction to generate single-stranded cDNAs (ss cDNA) that were tagged with specific sequences at each end. An oligo dT primer containing a specific sequence (CDS: AAGCAGTGGTAACAACGCAGAGTACT<sub>30</sub>VN (V=A,G,C N=A,G,C,T)) anneals at the polyA track at the 3' end of the mRNA and the reverse transcriptase (MMLV RnaseH<sup>-</sup>) transcribes the antisense strand until it reaches the end of the RNA strand when it adds additional C residues. If a primer (SMII: AAGCAGTGGTAACAACGCAGAGTACGCGGG or ML2G: AAGTGGCAACAGAGATAACGCGTACGCGGG) ending with 3 Gs is added, it anneals to the added Cs and the MMLV recognizes the rest of the primer sequence as template and continues transcription. As a result, the synthesized cDNAs contain specific sequence tags at both the 5' and the 3' end. When the 5' and the 3' ends are tagged with the same sequence (CDS and SMII) it is referred to as "symmetric". When the 5' end is tagged with a different sequence than the 3' end (CDS and ML2G) is referred to as "asymmetric". A double-stranded "cDNA library" is then generated by PCR amplification using the 3'PCR and ML2 primers (3' PCR: AAGCAGTGGTAACAACGCAGAGT and ML2: AAGTGGCAACAGAGATAACGCGT) that anneal to the added sequence tags.

The amplified "cDNA libraries" were manually arrayed onto nylon membranes with a 384 pin replicator. The DNA was denatured by alkali treatment, neutralized and cross-linked by UV light. The arrays were pre-hybridized with Express Hyb (Clontech) and hybridized with <sup>32</sup>P labeled probes generated by random hexamer priming of cDNA fragments corresponding to the



genes of interest. After washing, the blots were exposed to phosphorimaging cassettes and the intensity of the signal was quantified. The amount of the DNA on the arrays was also quantified by treating non-denatured or denatured arrays with Syber Green I or Syber Green II respectively (1:100,000 in 50mM Tris, pH8.0) for 2 minutes. After washing with 50mM Tris, pH8.0, the  
5 fluorescent emission was detected with a phosphorimager (Molecular Dynamics) and quantified. The amount of the arrayed DNA was used to normalize the hybridization signal and the corrected values are tabulated in Table 6.

Cell treatments. Several cell lines were treated with compounds to evaluate their effects on gene expression. There were eight treatments: 1) control, 2) low serum, 3) 200uM mimosine,  
10 4) 3mM HU, 5) 2uM AUR2 inhibitor, 6) 10uM cisplatin, 7) 400 ng/ml nocodazole-24 hours, and 8) 400 ng/ml nocodazole-48 hours. The treated cell lines are listed by cell line name followed by a number from 1 to 8.

#### Example 2c: Predicted proteins

15  
SGK187, CRIK (SEQ ID NO:1, ENCODING SEQ ID NO:58) encodes a protein that is 2053 amino acids long. It is classified as (superfamily/group/family): protein kinase, AGC, DMPK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position  
20 of the kinase catalytic region within the encoded protein is from amino acid 98 to amino acid 361. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 1970; percent identity = 96%; percent similarity = 98%; the accession number of the most similar entry in NRAA is AAC72823.1; the name or description, and species, of the  
25 most similar protein in NRAA is: Rho/rac-interacting citron kinase [Mus musculus]. Domains other than the kinase catalytic domain identified within this protein, and their amino acid positions, are: CNH domain (1620-1917); PH 1472-1591; Protein kinase C terminal domain (362-391); Phorbol esters/diacylglycerol binding domain (C1 domain) (1391-1439)".

SGK064, GRK7 (SEQ ID NO:2, ENCODING SEQ ID NO:59) encodes a protein that is  
30 553 amino acids long. It is classified as (superfamily/group/family): protein kinase, AGC, GRK.

The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 191 to amino acid 454. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 466; percent identity = 84%; percent similarity = 91%; the accession number of the most similar entry in NRAA is AAC95001.1; the name or description, and species, of the most similar protein in NRAA is: G protein-coupled receptor kinase GRK7 [*Spermophilus tridecemlineatus*]. Domains other than the kinase catalytic domain identified within this protein are: RGS (55-176).

10

SGK409, KIAA0303 (SEQ ID NO:3, ENCODING SEQ ID NO:60) encodes a protein that is 2523 amino acids long. It is classified as (superfamily/group/family): protein kinase, AGC, MAST. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 470 to amino acid 743 (PFAM profile). The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 2137; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is BAA20762.1; the name or description, and species, of the most similar protein in NRAA is: KIAA0303 [*Homo sapiens*]. Domains other than the kinase catalytic domain identified within this protein are: PDZ domain (1020-1148).

15

20

SGK021 (SEQ ID NO:4, ENCODING SEQ ID NO:61) encodes a protein that is 327 amino acids long. It is classified as (superfamily/group/family): protein kinase, AGC, Mo3C11.1\_ce. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 23 to amino acid 263. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 6.00E-138;

30

number of identical amino acids = 203; percent identity = 77%; percent similarity = 84%; the accession number of the most similar entry in NRAA is CAB76566.1; the name or description, and species, of the most similar protein in NRAA is: Serine/threonine protein kinase [Mus musculus] .

5

SGK410 (SEQ ID NO:5, ENCODING SEQ ID NO:62) encodes a protein that is 631 amino acids long. It is classified as (superfamily/group/family): protein kinase, AGC, PKC. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 289 to amino acid 557. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 555; percent identity = 95%; percent similarity = 96%; the accession number of the most similar entry in NRAA is NP\_002731.1 ; the name or description, and species, of the most similar protein in NRAA is: Protein kinase C, iota [Homo sapiens]. Domains other than the kinase catalytic domain identified within this protein are: Protein kinase C terminal domain (558-624); Phorbol esters/diacylglycerol binding domain (C1 domain) (176-225); Octicosapeptide repeat (100-129).

20

SGK069 (SEQ ID NO:6, ENCODING SEQ ID NO:63) encodes a protein that is 572 amino acids long. It is classified as (superfamily/group/family): protein kinase, AGC, Unique. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 158 to amino acid 421. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.10E-64; number of identical amino acids = 116; percent identity = 42%; percent similarity = 58%; the accession number of the most similar entry in NRAA is BAA36362.1; the name or description, and species, of the most similar protein in NRAA is: Serine/threonine protein kinase [Rattus norvegicus].

30

SGK110 (SEQ ID NO:7, ENCODING SEQ ID NO:64) encodes a protein that is 373 amino acids long. It is classified as (superfamily/group/family): protein kinase, AGC, Unique. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 68 to amino acid 329. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.10E-65; number of identical amino acids = 112; percent identity = 40%; percent similarity = 58%; the accession number of the most similar entry in NRAA is S71887; the name or description, and species, of the most similar protein in NRAA is: pk9.7 gastrula-specific PK [*Xenopus laevis*].

SGK053, CKLiK (SEQ ID NO:8, ENCODING SEQ ID NO:65) encodes a protein that is 357 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, AMPK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 23 to amino acid 279. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.60E-245; number of identical amino acids = 357; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP\_065130.1; the name or description, and species, of the most similar protein in NRAA is: CamKI-like protein kinase [*Homo sapiens*].

SGK124 (SEQ ID NO:9, ENCODING SEQ ID NO:66) encodes a protein that is 358 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, AMPK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 73 to amino acid 315. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 5.10E-252; number of identical amino acids = 358; percent identity = 100%; percent similarity = 100%; the accession number of

the most similar entry in NRAA is CAB81634.1; the name or description, and species, of the most similar protein in NRAA is: Novel protein kinase [Homo sapiens].

5 SGK254, CAMKKa (SEQ ID NO:10, ENCODING SEQ ID NO:67) encodes a protein that is 513 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, CAMK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 128 to amino acid 417. The results of a Smith Waterman search of the public database of amino acid  
10 sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.0e-323; number of identical amino acids = 470; percent identity = 91%; percent similarity = 94%; the accession number of the most similar entry in NRAA is A57156; the name or description, and species, of the most similar protein in NRAA is: Ca<sup>2+</sup>/calmodulin-dep. PK IV [Rattus norvegicus].

15

SGK297, CaMKIb2 (SEQ ID NO:11, ENCODING SEQ ID NO:68) encodes a protein that is 343 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, CAMK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The  
20 position of the kinase catalytic region within the encoded protein is from amino acid 15 to amino acid 270. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.40E-236; number of identical amino acids = 341; percent identity = 99%; percent similarity = 100%; the accession number of the most similar entry in NRAA is AAF74509.1; the name or description,  
25 and species, of the most similar protein in NRAA is: Ca<sup>2+</sup>/Calmodulin-dependent protein kinase I [Homo sapiens].

SGK411, CaMKII delta2 (SEQ ID NO:12, ENCODING SEQ ID NO:69) encodes a protein that is 499 amino acids long. It is classified as (superfamily/group/family): protein  
30 kinase, CAMK, CAMK. The kinase domain in this protein matches the hidden Markov profile

-269-

for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 14 to amino acid 272. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number  
5 of identical amino acids = 499; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is AAD20442.1; the name or description, and species, of the most similar protein in NRAA is: Multifunctional CAMK II delta2 [Homo sapiens].

10 SGK027 (SEQ ID NO:13, ENCODING SEQ ID NO:70) encodes a protein that is 436 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the  
15 kinase catalytic region within the encoded protein is from amino acid 74 to amino acid 325. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 7.70E-101; number of identical amino acids = 147; percent identity = 56%; percent similarity = 76%; the accession number of the most similar entry in NRAA is T22427; the name or description, and species, of the most  
20 similar protein in NRAA is: F49C5.4 - [Caenorhabditis elegans].

SGK046b (SEQ ID NO:14, ENCODING SEQ ID NO:71) encodes a protein that is 75 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 98 to profile position 175. The position of the  
25 kinase catalytic region within the encoded protein is from amino acid 1 to amino acid 75. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.40E-14; number of identical amino acids = 35; percent identity = 47%; percent similarity = 61%; the accession number of the most similar entry in NRAA is AAC33487.1; the name or description, and species, of the most  
30 similar protein in NRAA is: R31237\_1, partial CDS [Homo sapiens].

SGK046c (SEQ ID NO:15, ENCODING SEQ ID NO:72) encodes a protein that is 39 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 106 to profile position 144. The position of the kinase catalytic region within the encoded protein is from amino acid 1 to amino acid 39. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 6.40E-07; number of identical amino acids = 21; percent identity = 54%; percent similarity = 67%; the accession number of the most similar entry in NRAA is AAC33487.1; the name or description, and species, of the most similar protein in NRAA is: R31237\_1, partial CDS [Homo sapiens].

SGK089 (SEQ ID NO:16, ENCODING SEQ ID NO:73) encodes a protein that is 84 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 130 to profile position 216. The position of the kinase catalytic region within the encoded protein is from amino acid 1 to amino acid 84. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.40E-19; number of identical amino acids = 39; percent identity = 46%; percent similarity = 71%; the accession number of the most similar entry in NRAA is AAA97437.1; the name or description, and species, of the most similar protein in NRAA is: Serine/threonine kinase [Caenorhabditis elegans].

SGK133 (SEQ ID NO:17, ENCODING SEQ ID NO:74) encodes a protein that is 794 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 50 to amino acid 301. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 8.2e-318; number of identical

amino acids = 481; percent identity = 75%; percent similarity = 83%; the accession number of the most similar entry in NRAA is CAA07196.1; the name or description, and species, of the most similar protein in NRAA is: Putative serine/threonine protein kinase [Homo sapiens].

5           SGK004, MSK (SEQ ID NO:18, ENCODING SEQ ID NO:75) encodes a protein that is 786 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK-1. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 27 to amino acid  
10 281. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 784; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is P57059; the name or description, and species, of the most similar protein in NRAA is: SNF1LK [Homo sapiens].

15           SGK006 (SEQ ID NO:19, ENCODING SEQ ID NO:76) encodes a protein that is 262 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, EMK-1. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the  
20 kinase catalytic region within the encoded protein is from amino acid 25 to amino acid 261. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.50E-62; number of identical amino acids = 121; percent identity = 45%; percent similarity = 66%; the accession number of the most similar entry in NRAA is NP\_056570.1; the name or description, and species, of the  
25 most similar protein in NRAA is: Hormonally upregulated Neu-associated kinase [Mus musculus].

          SGK180, SNRK (SEQ ID NO:20, ENCODING SEQ ID NO:77) encodes a protein that is 765 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK,  
30 EMK-1. The kinase domain in this protein matches the hidden Markov profile for a full length



kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 16 to amino acid 269. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 757; percent identity = 99%; percent similarity = 99%; the accession number of the most similar entry in NRAA is AAF86944.1; the name or description, and species, of the most similar protein in NRAA is: HSNFRK [Homo sapiens].

SGK386, MLCKs (SEQ ID NO:21, ENCODING SEQ ID NO:78) encodes a protein that is 612 amino acids long. It is classified as (superfamily/group/family): protein kinase, CAMK, MLCK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 301 to amino acid 556. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 596; percent identity = 97%; percent similarity = 97%; the accession number of the most similar entry in NRAA is CAC10006.1; the name or description, and species, of the most similar protein in NRAA is: MYLK (myosin, light polypeptide kinase) [Homo sapiens].

SGK003 (SEQ ID NO:22, ENCODING SEQ ID NO:79) encodes a protein that is 337 amino acids long. It is classified as (superfamily/group/family): protein kinase, CKI, CKI. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 17 to amino acid 299. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.40E-215; number of identical amino acids = 304; percent identity = 91%; percent similarity = 95%; the accession number of the most similar entry in NRAA is NP\_001883.2; the name or description, and species, of the most similar protein in NRAA is: Casein kinase 1, alpha 1 [[Homo sapiens].

30

SGK066 (SEQ ID NO:23, ENCODING SEQ ID NO:80) encodes a protein that is 400 amino acids long. It is classified as (superfamily/group/family): protein kinase, CKI, CKI. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 21 to amino acid 281. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 5.40E-106; number of identical amino acids = 166; percent identity = 53%; percent similarity = 68%; the accession number of the most similar entry in NRAA is T24262; the name or description, and species, of the most similar protein in NRAA is: R90.1 [Caenorhabditis elegans].

SGK041, NKIAMRE (SEQ ID NO:24, ENCODING SEQ ID NO:81) encodes a protein that is 591 amino acids long. It is classified as (superfamily/group/family): protein kinase, CMGC, CDK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 4 to amino acid 286. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.0e-319; number of identical amino acids = 454; percent identity = 99%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP\_057592.1; the name or description, and species, of the most similar protein in NRAA is: NKIAMRE [Homo sapiens].

SGK112 (SEQ ID NO:25, ENCODING SEQ ID NO:82) encodes a protein that is 360 amino acids long. It is classified as (superfamily/group/family): protein kinase, CMGC, CDK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 4 to amino acid 304. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 8.70E-151; number of identical amino acids = 224; percent identity = 61%; percent similarity = 74%; the accession number of

the most similar entry in NRAA is NP\_004187.1; the name or description, and species, of the most similar protein in NRAA is: CDC2-related kinase [*Homo sapiens*].

SGK038, ERK7 (SEQ ID NO:26, ENCODING SEQ ID NO:83) encodes a protein that is  
5 371 amino acids long. It is classified as (superfamily/group/family): protein kinase, CMGC, MAPK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 13 to amino acid 323. The results of a Smith Waterman search of the public database of amino acid sequences  
10 (NRAA) with this protein sequence yielded the following results: Pscore = 1.80E-105; number of identical amino acids = 167; percent identity = 59%; percent similarity = 72%; the accession number of the most similar entry in NRAA is P51954; the name or description, and species, of the most similar protein in NRAA is: NEK1 (NIMA-RELATED PROTEIN KINASE 1) [*Mus musculus*].

15

SGK158 (SEQ ID NO:27, ENCODING SEQ ID NO:84) encodes a protein that is 751 amino acids long. It is classified as (superfamily/group/family): protein kinase, Microbial PK, ABC1. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 253 to profile position 261. The position  
20 of the kinase catalytic region within the encoded protein is from amino acid 522 to amino acid 532. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 3.30E-257; number of identical amino acids = 368; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP\_064632.1; the name or description,  
25 and species, of the most similar protein in NRAA is: Hypothetical protein [*Homo sapiens*].

SGK429 (SEQ ID NO:28, ENCODING SEQ ID NO:85) encodes a protein that is 626 amino acids long. It is classified as (superfamily/group/family): protein kinase, Microbial PK, ABC1. The kinase domain in this protein matches the hidden Markov profile for a full length  
30 kinase domain of 261 amino acids from profile position 1 to profile position 34. The position of

the kinase catalytic region within the encoded protein is from amino acid 200 to amino acid 315. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 6.00E-122; number of identical amino acids = 191; percent identity = 100%; percent similarity = 100%; the accession number of  
5 the most similar entry in NRAA is AAD43192.1; the name or description, and species, of the most similar protein in NRAA is: Putative human protein [Homo sapiens].

SGK152, SUDD (SEQ ID NO:29, ENCODING SEQ ID NO:86) encodes a protein that is 519 amino acids long. It is classified as (superfamily/group/family): protein kinase, Microbial  
10 PK, RI01. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 251 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 347 to amino acid 358. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical  
15 amino acids = 519; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP\_003822.1; the name or description, and species, of the most similar protein in NRAA is: sudD (suppressor of bimD6, Aspergillus nidulans) homolog [Homo sapiens].

20 SGK077 (SEQ ID NO:30, ENCODING SEQ ID NO:87) encodes a protein that is 798 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, C26C2\_ce. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 32. The position of the kinase catalytic region within the encoded protein is from amino acid 484 to  
25 amino acid 513. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 8.50E-306; number of identical amino acids = 492; percent identity = 62%; percent similarity = 73%; the accession number of the most similar entry in NRAA is BAB00640.1; the name or description, and species, of the most similar protein in NRAA is: Hapsin [Mus musculus].

30

SGK093, Wnk3 (SEQ ID NO:31, ENCODING SEQ ID NO:88) encodes a protein that is 1513 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, C26C2\_ce. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 162 to amino acid 420. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.10E-203; number of identical amino acids = 321; percent identity = 60%; percent similarity = 73%; the accession number of the most similar entry in NRAA is AAF31483.1; the name or description, and species, of the most similar protein in NRAA is: Kinase deficient protein KDP [Homo sapiens]. Domains other than the kinase catalytic domain identified within this protein are: Cyclin (amino acids 1373 to 1410).

SGK074 (SEQ ID NO:32, ENCODING SEQ ID NO:89) encodes a protein that is 355 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, DYRK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 6 to amino acid 342. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.00E-107; number of identical amino acids = 176; percent identity = 50%; percent similarity = 67%; the accession number of the most similar entry in NRAA is AAD41593.1; the name or description, and species, of the most similar protein in NRAA is: Myak-S [Mus musculus].

SGK087 (SEQ ID NO:33, ENCODING SEQ ID NO:90) encodes a protein that is 628 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, DYRK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 213 to amino acid 509. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA)

with this protein sequence yielded the following results: Pscore = 6.20E-183; number of identical amino acids = 267; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is AAF91393.1; the name or description, and species, of the most similar protein in NRAA is: DYRK4 [Homo sapiens].

5

SGK295, KIS (SEQ ID NO:34, ENCODING SEQ ID NO:91) encodes a protein that is 419 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, DYRK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 200. The position  
10 of the kinase catalytic region within the encoded protein is from amino acid 23 to amino acid 238. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.10E-292; number of identical amino acids = 415; percent identity = 99%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP\_058989.1; the name or description, and  
15 species, of the most similar protein in NRAA is: Kinase interacting with stathmin [Rattus norvegicus]. Domains other than the kinase catalytic domain identified within this protein are: RNA recognition motif (345-401).

SGK419 (SEQ ID NO:35, ENCODING SEQ ID NO:92) encodes a protein that is 661  
20 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, NAK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 51 to amino acid 341. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA)  
25 with this protein sequence yielded the following results: Pscore = 5.00E-252; number of identical amino acids = 355; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is CAB70863.1; the name or description, and species, of the most similar protein in NRAA is: Hypothetical protein [Homo sapiens].

SGK125, MYO3A (SEQ ID NO:36, ENCODING SEQ ID NO:93) encodes a protein that is 1615 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, NinaC. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 21 to amino acid 287. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 1615; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP\_059129.1; the name or description, and species, of the most similar protein in NRAA is: Myosin IIIA [Homo sapiens]. Domains other than the kinase catalytic domain identified within this protein are: Myosin head (340-1040); IQ, (3 domains, 1055-1366).

SGK445 (SEQ ID NO:37, ENCODING SEQ ID NO:94) encodes a protein that is 68 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, PLK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 71. The position of the kinase catalytic region within the encoded protein is from amino acid 1 to amino acid 68. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 9.90E-25; number of identical amino acids = 47; percent identity = 65%; percent similarity = 81%; the accession number of the most similar entry in NRAA is AAC37649.1; the name or description, and species, of the most similar protein in NRAA is: STK [Mus musculus].

SGK127 (SEQ ID NO:38, ENCODING SEQ ID NO:95) encodes a protein that is 945 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, RAF. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 661 to amino acid 922. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA)

with this protein sequence yielded the following results: Pscore = 8.60E-254; number of identical amino acids = 446; percent identity = 47%; percent similarity = 61%; the accession number of the most similar entry in NRAA is NP\_038599.1; the name or description, and species, of the most similar protein in NRAA is: Kinase suppressor of ras [Mus musculus]. Domains other than the kinase catalytic domain identified within this protein are: Phorbol esters/diacylglycerol binding domain (C1 domain) (408-451).

SGK009, ANKRD3 (SEQ ID NO:39, ENCODING SEQ ID NO:96) encodes a protein that is 832 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, RIP. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 22 to amino acid 276. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 784; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP\_065690.1; the name or description, and species, of the most similar protein in NRAA is: Ankyrin repeat domain 3 [Homo sapiens]. Domains other than the kinase catalytic domain identified within this protein are: "Ankyrin (10 domains: 437-469, 470-502, 503-535, 536-568, 569-602, 602-635, 636-668, 669-701, 702-730, 738-770)".

SGK421, STK22A, TSK1 (SEQ ID NO:40, ENCODING SEQ ID NO:97) encodes a protein that is 367 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, STK22A. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 12 to amino acid 272. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 6.40E-209; number of identical amino acids = 307; percent identity = 84%; percent similarity = 90%; the accession number of the most similar entry in NRAA is NP\_033461.1; the name or description,



and species, of the most similar protein in NRAA is: STK 22A (spermiogenesis associated) [Mus musculus].

5 SGK047 (SEQ ID NO:41, ENCODING SEQ ID NO:98) encodes a protein that is 31 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, STKR. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 232 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 1 to amino acid 31. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) 10 with this protein sequence yielded the following results: Pscore = 0.265466; number of identical amino acids = 11; percent identity = 46%; percent similarity = 67%; the accession number of the most similar entry in NRAA is NP\_004320.1; the name or description, and species, of the most similar protein in NRAA is: Bone morphogenetic protein receptor, type IA [Homo sapiens].

15 SGK196 (SEQ ID NO:42, ENCODING SEQ ID NO:99) encodes a protein that is 350 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, Unique. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 88. The position of the kinase catalytic region within the encoded protein is from amino acid 81 to amino acid 160. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) 20 with this protein sequence yielded the following results: Pscore = 2.70E-248; number of identical amino acids = 350; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is BAB15623.1; the name or description, and species, of the most similar protein in NRAA is: Unnamed protein product [Homo sapiens].

25 SGK396 (SEQ ID NO:43, ENCODING SEQ ID NO:100) encodes a protein that is 472 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, Unique. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 186 to amino acid 425. 30

The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 3.00E-09; number of identical amino acids = 34; percent identity = 28%; percent similarity = 50%; the accession number of the most similar entry in NRAA is BAB11570.1; the name or description, and species, of the most similar protein in NRAA is: Receptor-like protein kinase [Arabidopsis thaliana].

SGK279, PKN (SEQ ID NO:44, ENCODING SEQ ID NO:101) encodes a protein that is 424 amino acids long. It is classified as (superfamily/group/family): protein kinase, Other, YWY3\_ce. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 53 to amino acid 313. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 4.10E-276; number of identical amino acids = 400; percent identity = 94%; percent similarity = 95%; the accession number of the most similar entry in NRAA is BAA36362.1; the name or description, and species, of the most similar protein in NRAA is: Serine/threonine protein kinase [Rattus norvegicus].

SGK037 (SEQ ID NO:45, ENCODING SEQ ID NO:102) encodes a protein that is 649 amino acids long. It is classified as (superfamily/group/family): protein kinase, STE, NEK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 4 to amino acid 259. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 4.10E-56; number of identical amino acids = 176; percent identity = 65%; percent similarity = 83%; the accession number of the most similar entry in NRAA is P51954; the name or description, and species, of the most similar protein in NRAA is: NEK1 [Mus musculus].

SGK060 (SEQ ID NO:46, ENCODING SEQ ID NO:103) encodes a protein that is 645 amino acids long. It is classified as (superfamily/group/family): protein kinase, STE, NEK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 29 to amino acid 287. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.3e-322; number of identical amino acids = 463; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is BAB15672.1; the name or description, and species, of the most similar protein in NRAA is: Unnamed protein product [*Homo sapiens*].

SGK080 (SEQ ID NO:47, ENCODING SEQ ID NO:104) encodes a protein that is 446 amino acids long. It is classified as (superfamily/group/family): protein kinase, STE, NEK. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 8 to amino acid 269. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 3.80E-266; number of identical amino acids = 404; percent identity = 90%; percent similarity = 93%; the accession number of the most similar entry in NRAA is NP\_002488.1; the name or description, and species, of the most similar protein in NRAA is: NIMA (never in mitosis gene a)-related kinase [*Homo sapiens*].

SGK002 (SEQ ID NO:48, ENCODING SEQ ID NO:105) encodes a protein that is 399 amino acids long. It is classified as (superfamily/group/family): protein kinase, STE, STE11. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 71 to amino acid 368. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 3.10E-248; number of identical amino acids = 369; percent identity = 92%; percent similarity = 94%; the accession number of

the most similar entry in NRAA is P36507; the name or description, and species, of the most similar protein in NRAA is: MEK2 [Homo sapiens].

5       SGK058 (SEQ ID NO:49, ENCODING SEQ ID NO:106) encodes a protein that is 278 amino acids long. It is classified as (superfamily/group/family): protein kinase, STE, STE11. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 11 to amino acid 274. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA)  
10       with this protein sequence yielded the following results: Pscore = 3.00E-116; number of identical amino acids = 167; percent identity = 99%; percent similarity = 99%; the accession number of the most similar entry in NRAA is BAB15538.1; the name or description, and species, of the most similar protein in NRAA is: Unnamed protein product [Homo sapiens].

15       SGK103 (SEQ ID NO:50, ENCODING SEQ ID NO:107) encodes a protein that is 28 amino acids long. It is classified as (superfamily/group/family): protein kinase, STE, STE11. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 125 to profile position 148. The position of the kinase catalytic region within the encoded protein is from amino acid 1 to amino acid 24. The  
20       results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0.008942; number of identical amino acids = 16; percent identity = 57%; percent similarity = 68%; the accession number of the most similar entry in NRAA is CAA39285.1; the name or description, and species, of the most similar protein in NRAA is: Fused [Drosophila melanogaster].

25       SGK035 (SEQ ID NO:51, ENCODING SEQ ID NO:108) encodes a protein that is 348 amino acids long. It is classified as (superfamily/group/family): protein kinase, STE, STE20. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the  
30       kinase catalytic region within the encoded protein is from amino acid 73 to amino acid 324. The

results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 7.60E-212; number of identical amino acids = 318; percent identity = 92%; percent similarity = 95%; the accession number of the most similar entry in NRAA is Q13177; the name or description, and species, of the most similar protein in NRAA is: PAK-2 [Homo sapiens].

SGK075 (SEQ ID NO:52, ENCODING SEQ ID NO:109) encodes a protein that is 1106 amino acids long. It is classified as (superfamily/group/family): protein kinase, STE, STE20. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 15 to amino acid 281. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 4.00E-158; number of identical amino acids = 227; percent identity = 67%; percent similarity = 80%; the accession number of the most similar entry in NRAA is NP\_059129.1; the name or description, and species, of the most similar protein in NRAA is: Myosin IIIA [Homo sapiens].

SGK188, EphA9 (SEQ ID NO:53, ENCODING SEQ ID NO:110) encodes a protein that is 1009 amino acids long. It is classified as (superfamily/group/family): protein kinase, TK, RTK-11. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 645 to amino acid 899. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 538; percent identity = 54%; percent similarity = 71%; the accession number of the most similar entry in NRAA is Q61772; the name or description, and species, of the most similar protein in NRAA is: Ephrin Type-A receptor 7 [Mus musculus]. Domains other than the kinase catalytic domain identified within this protein are: "Ephrin receptor ligand binding domain (35-211); Fibronectin type III domain (2 domains: 339-436 & 454-537);

SAM domain (Sterile alpha motif)(931-996)".

SGK040 (SEQ ID NO:54, ENCODING SEQ ID NO:111) encodes a protein that is 909 amino acids long. It is classified as (superfamily/group/family): protein kinase, TK, Unique.

5 The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 394 to amino acid 646. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 4.50E-26; number of identical

10 amino acids = 76; percent identity = 30%; percent similarity = 52%; the accession number of the most similar entry in NRAA is P41243; the name or description, and species, of the most similar protein in NRAA is: BATK [Rattus norvegicus].

SGK390 (SEQ ID NO:55, ENCODING SEQ ID NO:112) encodes a protein that is 1164 amino acids long. It is classified as (superfamily/group/family): protein kinase -like, DAG kin, DAG kin. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 1 to profile position 13. The position of the kinase catalytic region within the encoded protein is from amino acid 383 to amino acid 395. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino

20 acids = 1063; percent identity = 91%; percent similarity = 94%; the accession number of the most similar entry in NRAA is Q64398; the name or description, and species, of the most similar protein in NRAA is: Diacylglycerol kinase eta [Mesocricetus auratus]. Domains other than the kinase catalytic domain identified within this protein are: "Diacylglycerol kinase catalytic

25 domain (332-457);  
DAG\_PE-bind (176-225 & 248-298); PH domain (66-158)  
Diacylglycerol kinase accessory domain (770-927);".

SGK007 (SEQ ID NO:56, ENCODING SEQ ID NO:113) encodes a protein that is 884 amino acids long. It is classified as (superfamily/group/family): protein kinase -like, GCyc,

30

GCyc. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 166 to profile position 261. The position of the kinase catalytic region within the encoded protein is from amino acid 613 to amino acid 716. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 5.20E-265; number of identical amino acids = 499; percent identity = 51%; percent similarity = 62%; the accession number of the most similar entry in NRAA is T42260; the name or description, and species, of the most similar protein in NRAA is: Guanylate cyclase [*Rattus norvegicus*].

SGK050 (SEQ ID NO:57, ENCODING SEQ ID NO:114) encodes a protein that is 48 amino acids long. It is classified as (superfamily/group/family): protein kinase -like, GCyc, GCyc. The kinase domain in this protein matches the hidden Markov profile for a full length kinase domain of 261 amino acids from profile position 51 to profile position 89. The position of the kinase catalytic region within the encoded protein is from amino acid 1 to amino acid 37. The results of a Smith Waterman search of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.30E-10; number of identical amino acids = 48; percent identity = 23%; percent similarity = 48%; the accession number of the most similar entry in NRAA is NP\_003986.2; the name or description, and species, of the most similar protein in NRAA is: Natriuretic peptide receptor B precursor, isoform b [*Homo sapiens*].

## Results

The results of the microarray expression analysis of the protein kinases presented in this application are shown, for example, in Tables 6 and 7. In Table 6, data presentation from left to right is as follows: "Tissue": tissue type of the cDNA; "Tumor sym", indicates that the tissue is derived from a tumor, "sym" refers to the fact that the 5' and 3' primers used to make the sample are the same; "Normal Sym", indicates normal tissue was used to make the sample, with symmetric primers as described above; "Tumor 1o", indicates that primary tumor tissue was used to make Th cDNA; "Tumor cells", indicates that these cDNA samples were made from cultured tumor cells; "Normal", indicates that these samples are derived from normal tissue or cell lines; "Endos", indicates that these samples are derived from endothelium-related tissue sources; "p53"

refers to the status, mutant or wild-type, of the p53 gene in the source samples. Normalized expression values are presented for each gene referred to by its SEQ\_ID# on the subsequent columns.

Genes represented in the tissue array are listed below: SGK187 (SEQ ID NO:1); SGK124  
 5 (SEQ ID NO:9); SGK386 (SEQ ID NO:21); SGK003 (SEQ ID NO:22); SGK077 (SEQ ID NO:30); SGK093 (SEQ ID NO:31); SGK074 (SEQ ID NO:32); and SGK396 (SEQ ID NO:43), as shown in Table 6

SGK187 is expressed highest in the following tissues: heart, SA-OS cells, Prostate Tumor\_13, fetal liver, C33A, Prostate Tumor10. This expression pattern suggests this kinase  
 10 may play a role in heart-related pathologies, and/or prostate cancer.

SGK074 is expressed highest in the following tissues: OVCAR-4 - 2, HCT-116 - 1, OVCAR-5 - 7, Prostate Tumor - 09, Prostate Tumor - 08, U2OS - 1, adrenal gland - h, Prostate Tumor- 18, HT29 - 5, SW480 - 3. This expression pattern suggests expression of this gene may be associated with ovarian cancer, prostate cancer, colon cancer, and other forms of neoplasia.

15 Sgk093 is expressed highest in the following tissues: fetal kidney, kidney, prostate tumor, fetal liver, Caki-1, U2OS - 6, ACHN, normal prostate. This expression pattern suggests that this gene may be involved in diseases of the prostate, kidney and/or liver.

SGK396 is expressed highest in the following tissues: cerebellum, Prostate Tumor18, fetal brain, thymus -h, and lymph node. This expression pattern suggests this gene may be  
 20 involved in hematopoietic disease, prostate cancer, or CNS abnormalities involving the cerebellum.

A statistical analysis of the gene expression patterns was carried out from the tissue array data. The tissue array data for the 8 kinases were standardized for statistical analysis across the different tissue types using range standardization. Standardization converts measurements to a  
 25 common scale. We used range standardization, which subtracts the smallest value of each variable from each value and divides by its range. The new scale starts at 0 and ends at 1.0. The following statistical procedures were implemented on the standardized data: generation of descriptive statistics, graphical visualization, hierarchical and k-means cluster analysis, and comparison of groups using analysis of variance (ANOVA). When data was present for both  
 30 normal and tumor tissues for particular tissue types, such groups were directly compared for fold



differences. All statistical analyses were carried out separately for the symmetric and asymmetric tissue array laboratory methods (we know from experience with past data that gene expression is dependent upon the method used). All statistical analysis was carried out using SYSTAT 9.01 (Copyright © 1999 by SPSS, Inc.).

5

## SUMMARY OF RESULTS:

### 1. SGK187 (NA#1)

#### 10 (Symmetric data)

- Expressed highest in normal tissue of the human heart, fetal liver, and thymus.
- Next highest expressions in the following tumor types: ovary, renal, lung, melanoma, colon, neuro.
- 15 • Expressed in tumor but not normal tissue of the following tissue types: lung, prostate, renal.
- 1.4-fold higher expression in tumor vs. normal colon tissue.
- A k-means cluster shows association of this kinase in renal tumor and ovary adenocarcinoma.

#### (Asymmetric data)

20

- Highest expression in tumors associated with the following tissue types: neuro, endo, ovary, breast, and colon.
- Expressed in tumor but not normal sample of breast, colon, and renal tissue.
- Over-expressed (tumor vs. normal) in the following tissues: neuro (3.4x) and pancreas (3.5x).
- 25 • Under-expressed (tumor vs. normal) in the following tissues: lung (2.0x) and prostate (4.1x).
- Singleton and tumor-associated small clusters (k-means) for this kinase were formed by neuro-metastases sample singly and in association with endo, neuro, and ovary tumors.

### 2. SGK124 (NA#9)

## (Symmetric data)

- Highest expression in leukemia.
- 5 • Next highest expressions in the following tissues: (1) normal: human keratinocy, liver, pituitary gland, fetal liver, lymph node, bone marrow, and fetal lung; and (2) tumor: melanoma, renal, neuro, lung, ovary, colon, and breast.
- Expressed in tumor but not normal sample of breast tissue.
- Over-expressed (tumor vs. normal) in the following tissues: colon (1.5x) and renal (5.5x).
- 10 • Under-expressed (tumor vs. normal) in the following tissues: lung (2.8x) and neuro (3.2x); and expressed in normal but not tumor sample of testes.
- A leukemia sample and a melanoma sample each form a singleton cluster (k-means) for this kinase.

## 15 (Asymmetric data)

- Highest expression in tumors associated with the following tissue types: neuro, endo, ovary, breast, colon, renal, and lung.
- Expressed in tumor but not normal sample of breast and colon tissue.
- 20 • Over-expressed (tumor vs. normal) in the following tissues: neuro (4.0x) and prostate (2.2x).
- Under-expressed (tumor vs. normal) in the following tissues: endo (2.0x) and pancreas (5.3x).
- Singleton clusters (k-means) for this kinase were formed by a bone tumor sample (h-osteogenic) and a neuro (metastases) sample. Other distinct tumor- associated clusters
- 25 included one with cervical tumor and neuro-metastases, colon and lung tumor, and an association of adrenal gland with a metastases tumor sample.

## 3. SGK386 (NA#21)

(Symmetric data)

- Only expressed in a colon tumor sample. Zero expression for all other samples.

5 (Asymmetric data)

- Highest expression in tumors associated with the following tissue types: prostate metastases samples, endo, lung, ovary, breast, and colon.
- Expressed in tumor but not normal sample of breast, colon, endo, and neuro.
- 10 • Singleton and tumor-associated small clusters (k-means) for this kinase were formed by metastases samples, primarily prostate and neuro.

4. SGK003 (NA#22)

15 (Symmetric data)

- Highest expression in tumors of the colon and breast. Also high expression in normal tissue of the adrenal gland.
- Over-expressed (tumor vs. normal) in the following tissues: colon (2.7x), breast (4.5x), and  
20 testes (4.6x).
- Under-expressed (tumor vs. normal) in the following tissues: lung (2.1x), neuro (2.8x), prostate (10.1x), and renal (3.6x).
- A colon adenocarcinoma sample forms a singleton cluster (k-means) for this kinase. A breast tumor sample forms another cluster in association with adrenal gland, both of which are high  
25 expressors for this kinase.

(Asymmetric data)

- Highest expression in tumors associated with the following tissue types: colon, lung, metastases-prostate, ovary, and bone.
- Expressed in tumor but not normal sample of breast.
- Over-expressed (tumor vs. normal) of the following tissues: colon (887x), endo (73x), neuro (75x), and renal (15x).
- No under-expression (tumor vs. normal).
- Singleton and tumor-associated small clusters (k-means) for this kinase were formed by prostate metastases samples with colon and lung tumors.

10 5. SGK077 (NA#30)

(Symmetric data)

- Highest expression in normal samples of human lung, cerebellum, pituitary gland, lymph node, fetal brain, thymus, pancreas, and placenta. Also relatively high expression in the following tumor types: neuro, breast, renal, colon, and melanoma.
- Over-expressed (tumor vs. normal) in the following tissues: prostate (6.3x) and breast (3.7x).
- Under-expressed (tumor vs. normal) in the following tissues: colon (1.7x), lung (22.9x), neuro (3.7x).

20

(Asymmetric data)

- Highest expression in metastases-prostate sample, normal human stomach, and tumor samples of endo, colon, neuro, lung, and ovary.
- Over-expressed (tumor vs. normal) in neuro tissue (6.9x).
- Under-expressed (tumor vs. normal) in the following tissue types: endo (2.0x), lung (3.1x), pancreas (3.7x), and prostate (4.0x).
- Most k-means clusters were formed by HUVEC normal samples, some of these in association with tumor samples.

## 6. SGK093 (NA#31)

(Symmetric data)

5

- Highest expression in normal human adult and fetal kidney, fetal liver, and tumor samples of renal and lung.
- Expressed in tumor but not normal sample of prostate tissue.
- Over-expressed (tumor vs. normal) of breast tissue (1.4x).
- 10 • Under-expressed (tumor vs. normal) in the following tissues: colon (2.0x), lung (1.9x), neuro (2.2x), and renal (15.9x).
- Distinct k-means clusters were observed for the following singletons or associations: renal clear cell carcinoma (singleton), renal tumor (singleton), and fetal lung with lung carcinoma and lung adenocarcinoma (association).

15

(Asymmetric data)

- Highest expression in metastases samples and other tumors of bone, renal, ovary, and lung.
- Expressed in tumor but not normal samples of colon and renal tissue.
- 20 • Over-expressed (tumor vs. normal) in breast (2.3x) and lung (1.7x).
- Under-expressed (tumor vs. normal) in the following tissue types: endo (4.0x), neuro (2.2x), and prostate (124x).
- Singleton or very small k-means clusters were formed by metastases samples and one cluster was formed by lung adenocarcinoma in association with ovarian cancer.

25

## 7. SGK074 (NA#32)

(Symmetric data)

- Highest expression in normal adrenal gland and tumor samples of colon and breast.
- Over-expressed (tumor vs. normal) in the following tissues: colon (1.4x), prostate (1.9x), and breast (2.3x).
- Under-expressed (tumor vs. normal) in the following tissues: lung (2.1x), neuro (2.6x), renal (3.2x), and testes (2.1x).
- Distinct k-means clusters were observed for the following singletons: colon adenocarcinoma and breast adenocarcinoma.

(Asymmetric data)

10

- Highest expression in a large variety of tumors including ovary, colon, metastases samples of prostate, bone, lung, neuro, cervical, breast, and endo.
- Generally, very highly expressed in tumor as versus normal samples of a wide variety of tissues. Over-expressed (tumor vs. normal) in the following tissue types: breast (6.1x), colon (16.9x), endo (5.7x), neuro (9.1x), and pancreas (1.8x).
- Under-expressed (tumor vs. normal) in the following tissue types: lung (1.5x), prostate (1.7x), and renal (2.7x).
- Singleton or very small k-means clusters were formed by ovarian tumor and metastases samples in association with one another, and also in association with other tumor types.

20

#### 8. SGK396 (NA#43)

(Symmetric data)

- Highest expression in the following normal tissues: cerebellum, fetal brain, thymus, lymph node, pituitary gland, fetal lung, and fetal kidney. Also very highly expressed in lung tumor samples and a host of other tumor samples: melanoma, renal, ovarian, leukemia, neuro, breast, and prostate.
- Expressed in tumor but not normal samples of prostate and breast.

- Over-expressed (tumor vs. normal) in testes (1.5x).
- Under-expressed (tumor vs. normal) in the following tissues: colon (1.5x), lung (1.5x), neuro (8.1x), and renal (3.9x).
- Two cancer dominated k-means clusters were observed, which represented a wide variety of tumor types.

(Asymmetric data)

- Highest expression in a large variety of tumors including prostate and neuro metastases samples, and tumors of bone, ovary, endo, breast, colon, and lung.
- Over-expressed (tumor vs. normal) in the following tissue types: colon (6.3x) and neuro (4.0x)
- Under-expressed (tumor vs. normal) in the following tissue types: endo (1.5x), lung (1.7x), pancreas (1.7x), prostate (4.2x), and renal (1.7x).
- Singleton or very small k-means clusters were formed by different metastases samples from different tissue types, and in association with other tumor types.

#### PCR Screening:

#### Screening for expression sources by PCR from ds cDNA templates

##### Preparation of dscDNA templates

dscDNA templates were prepared by PCR amplification of symmetrically-tagged reverse transcriptase sscDNA products generated as described in detail under Materials and Methods for the Tissue Array Gene Expression protocol. The tissue sources amplified are listed, for example, in Table 7. The amplification conditions were as follows: per 200 microl of PCR reaction, added 100 microl of Premix TaKaRa ExTaq, 20.0 microl of pwo DNA polymerase (1/10 dilution made as follows: 1 microl pwo (5 units/microl), 1 microl 10x PCR buffer with 20 mM MgSO<sub>4</sub>, 8 microl water), 4.0 microl sscDNA template (reverse transcriptase product), 8.0 microl 10 pmoles/microl (10 microM) primer (AAGCAGTGGTAACAACGCAGAGT ) (1.0 microM final

conc.) and 68.0 microl H<sub>2</sub>O. The reaction was amplified according to the following regiment: hot start (95°C for 1 min), 95°C for 1 min, 24 cycles, 95°C for 20 s, 65°C for 30 s, 68°C for 6 min, 68°C for 10 min, 1 cycle and 4°C forever. Following the PCR reaction, 5-10 microl of product were applied to an agarose gel together with 1kb ladder size standards to assess the yield and uniformity of the product.

### Source identification

The following relates to Table 7. 5 microl of each of the 48 sources were used as a template with primers designed from the nucleotide sequence of the genes whose expression pattern were to be determined. The PCR conditions were the same as those used to make the dscDNA templates except for the following modifications: the PCR reaction was scaled down from 100 to 20 microl; pwo was omitted; 5 microl of template were used. The cycling conditions were the same as those used to make the dscDNA templates except that 35 instead of 24 cycles were used in the protocol. The primers used in these amplifications are listed below.

SGK069 (SEQ ID NO:6) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GGACATGATGACGCTGAGTGCTCAG; 3' primer: GTGCTCCCTGATGGCGATCACAGCG. Primer sequences were designed from partial or full-length versions of SEQ ID NO:6.

SGK110 (SEQ ID NO:7) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: CTCCTGTCTTCTTGACCTCAGGGC; 3' primer: CAGGTCCAGAAGCCCCTGGAGC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:7.

SGK053 (SEQ ID NO:8) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: CCGAAAATCTCTTGACTACAGTC; 3' primer: CATAGCCTGGAGTTCCACAGGCAG. Primer sequences were designed from partial or full-length versions of SEQ ID NO:8.



SGK254 (SEQ ID NO:10) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GAGCCTACTAGAAACGGTGTGGACC; 3' primer: CTTACCAAACCCTTCTTTCACCAG. Primer sequences were designed from partial  
5 or full-length versions of SEQ ID NO:10.

SGK411 (SEQ ID NO:12) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: AAAGGCCAGAGGGGACAGACGTGG; 3' primer: CCTCAATTGTTGTATTTGAACTCTC. Primer sequences were designed from partial or full-  
10 length versions of SEQ ID NO:12.

SGK027 (SEQ ID NO:13) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GAAAAGGTGGCCATTAAGATCCTGG; 3' primer: CATCCATTCATTCATGATGCAG. Primer sequences were designed from partial or full-  
15 length versions of SEQ ID NO:13.

SGK046b (SEQ ID NO:14) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GAAGAGGAGGAGGCCAGACCATG; 3' primer: GGAAGAGTTCTGTGGCCACATAGG. Primer sequences were designed from partial or full-  
20 length versions of SEQ ID NO:14.

SGK089 (SEQ ID NO:16) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: CTTCTTTGACAAAGAACGTAACATC; 3' primer: CAGTATATCCTTACAGTTCTGCTC. Primer sequences were designed from partial or full-  
25 length versions of SEQ ID NO:16.

SGK003 (SEQ ID NO:22) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: CAACAGCGGCTCCAAAGCCGAACTC. 3' primer: TGCCTGTCTGGGTTTGGGCCTGCTG. Primer sequences were designed from partial or full-  
30 length versions of SEQ ID NO:22.

SGK066 (SEQ ID NO:23) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GTCCCGAGGCACATTCACCATTAG. 3' primer: GATGTCTCGATGCAAGAATCCCAC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:23.

SGK041 (SEQ ID NO:24) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GAGATGTATGAAACCCTTGGAAG. 3' primer: CATAAGACACTATGAACACGATTC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:24.

SGK038 (SEQ ID NO:26) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GCCTATGGCATTGTGTGGAAGGCAG. 3' primer: CAGCAGGGGCAGGTGGCCAGTGCC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:26.

SGK429 (SEQ ID NO:28) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GTTTCGGTTTCACCGCAGCCTCGCC. 3' primer: TGGAGAAGGAAAGAATGGGAGAAAC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:28.

SGK093 (SEQ ID NO:31) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GAAAAGGAGGACATGGAGACCCAGG. 3' primer: CTTTACGCAGCTTCTCACGCTTTCG. Primer sequences were designed from partial or full-length versions of SEQ ID NO:31.

SGK087 (SEQ ID NO:33) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: CCTAAAGCTTTTAAAGAACCAGCTG. 3' primer: CTGATGAATCCAAGCATGCTTGAGG. Primer sequences were designed from partial or full-length versions of SEQ ID NO:33.

SGK009 (SEQ ID NO:39) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GATTTTGGTCTGGCCAAGTGCAACG. 3' primer: GTATACATCGTGCTTGGTGTCGAAG. Primer sequences were designed from partial or full-length versions of SEQ ID NO:39.

SGK421 (SEQ ID NO:40) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: AGGGCTCCTATGCAAAAGTAAAATC. 3' primer: CCAGCAGTGGCTGAGGATCTCG. Primer sequences were designed from partial or full-length versions of SEQ ID NO:40.

SGK396 (SEQ ID NO:43) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GGAACGAGATCTTCTTGATGCTGAG. 3' primer: GGAAGGGCTGCTGCGAAGCTTTCTG. Primer sequences were designed from partial or full-length versions of SEQ ID NO:43.

SGK037 (SEQ ID NO:45) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GAATGGCAGGCTGTTTATTGTAATG. 3' primer: GGAATAAGATTCTCTAAAAAGGGC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:45.

SGK060 (SEQ ID NO:46) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: CAGCTAAGTGTGTGAGTGGATCAAC; 3' primer: CCATCGTGATCAGCAACTGCTCTTC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:46.

SGK080 (SEQ ID NO:47) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: CTGAGAACTATGAAGTGTGTACAC. 3' primer: CCAGGATCTGTCTGCTTTTGTAGTTG. Primer sequences were designed from partial or full-length versions of SEQ ID NO:47.

SGK002 (SEQ ID NO:48) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GCCGGCGCTCACCATCAACCCTACC. 3' primer: CTCACTGAGCATCTTTAGGTCCGCC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:48.

SGK058 (SEQ ID NO:49) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: CCTATCCTATGGACCAAGGGTGAG. 3' primer: GTGCTCCGATGTAAAACATGGCGG. Primer sequences were designed from partial or full-length versions of SEQ ID NO:49.

SGK035 (SEQ ID NO:51) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GATGATGAAGAGACTGCCCTCCC. 3' primer: CATTGCTTCTTTAGCTGCCATGATC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:51.

SGK075 (SEQ ID NO:52) was amplified from the 48 cDNA panel source using the following oligonucleotide primers: 5' primer: GATGCCATGCAGGGCCTTCAGC. 3' primer: CAAAGTCAACGAGCTTAACTCCTC. Primer sequences were designed from partial or full-length versions of SEQ ID NO:52.

#### Multiple Tissue Expression blots (MTE)

MTE (Multiple Tissue Expression) blots were obtained from Clontech Laboratories, Inc. These blots contained 84 arrayed cDNA samples derived from normal human tissue and human cell lines, and controls. The expression blots were prehybridized with ExpressHyb hybridization solution (Clontech Laboratories) containing 0.1 mg/ml denatured salmon sperm DNA at a temperature of 65 °C for two hours. Radioactive DNA probes were prepared using the Random Priming DNA labeling kit (Roche). Generation of DNA probe for SGK093 (Wnk3) Two synthetic oligonucleotides were designed which will amplify a 538 base pair fragment of Wnk3.

The upstream oligonucleotide has the sequence 5' CCGCCACGCCAGCTACTCATCTAC 3'.  
The downstream oligonucleotide has the sequence 5' TCTCTTCAACAGGGTCTCCACTCG 3'.  
A fragment of the expected size was amplified by polymerase chain reaction from a human testes  
cDNA library and the sequence confirmed by DNA sequencing. Generation of DNA probe for  
5 SGK188 (EphA9): Two synthetic oligonucleotides were designed which will amplify a 387 base  
pair fragment of EphA9. The upstream oligonucleotide has the sequence 5'  
TTCTCCCAGATCCACAGC 3'. The downstream oligonucleotide has the sequence 5'  
ACCGGAGTCCTGCCTTGGAAG 3'. A fragment of the expected size was amplified by  
polymerase chain reaction from a human testes cDNA library and the sequence confirmed by  
10 DNA sequencing. Purified DNA fragments (100 ng) were labeled with 250 uCi of 32P-labeled  
dCTP for 45 minutes using the kit protocol. Unincorporated nucleotide was removed through  
the use of a spin column (ProbeQuant G50 micro columns, Amersham Pharmacia, Inc.). After  
denaturation by boiling for three minutes, the probe was introduced into the prehybridization  
solution, and the blot was hybridized at 65 °C for 20 hours. The blot was subsequently washed  
15 four times for 15 minutes each at 65 °C in a solution containing 15 mM NaCl, 1.5 mM  
Na<sub>3</sub>Citrate, 0.1% sodium lauryl sulfate (SDS) and exposed to the phosphoimager screen for  
quantitation.

#### Northern blot (SGK188)

#### 20 Materials and Methods for the Northern blot

The northern blot was obtained from Clontech Laboratories, Inc. This blot contained 12  
poly A+ RNA samples from human tissues in the following order: whole brain, heart, skeletal  
muscle, colon, thymus, spleen, kidney, liver, small intestine, placenta, lung, peripheral blood  
leukocyte. The expression blots were prehybridized with ExpressHyb hybridization solution  
25 (Clontech Laboratories) containing 0.1 mg/ml denatured salmon sperm DNA at a temperature of  
65 °C for two hours. A radioactive DNA probe was prepared using the Random Priming DNA  
labeling kit (Roche). Purified DNA fragment (100 ng) was labeled with 250 uCi of 32P-labeled  
dCTP for 45 minutes using the kit protocol. Unincorporated nucleotide was removed through the  
use of a spin column (ProbeQuant G50 micro columns, Amersham Pharmacia, Inc.). After  
30 denaturation by boiling for three minutes, the probe was introduced into the prehybridization

solution, and the blot was hybridized at 65 °C for 20 hours. The blot was subsequently washed four times for 15 minutes each at 65 °C in a solution containing 15 mM NaCl, 1.5 mM Na<sub>3</sub>Citrate, 0.1% sodium lauryl sulfate (SDS) and exposed to the phosphoimager screen for quantitation.

- 5 Analysis of the northern blot shows expression of SGK188 in human brain and colon. Two transcripts are detected of approximately 4.4 kb and 6 kb in size.

**EXAMPLE 2e: Classification of polypeptides exhibiting kinase activity among defined groups**

10

**AGC Group**

Potential biological and clinical implications of the novel AGC group protein kinases are described next. SGK187 (SEQ ID NO:58) is a new DMPK family member that lists as its prototype myotonic dystrophy protein kinase (DMPK). Since the initial filing of this application, the murine orthologue of human SGK187 appeared in the public database as the full-length gene encoding Crik (AF086824) (J. Biol. Chem. 273 (45), 29706-29711 (1998)). Motif analysis of the amino acid sequence predicted for SGK187 (SEQ ID#58) showed the presence of four extracatalytic C-terminal domains: CNH, a plekstrin homology (PH) domain, a protein kinase C terminal domain and a phorbol ester/diacyl glycerol binding (C1) domain. Murine Crik is made as two different isoforms, a long 240-kDa protein in which the kinase domain is followed by the sequence for Citron, a Rho/Rac-binding protein comprising the CNH, PH and C1 domains. The second form of Crik is a short 54-kDa protein (CRIK-short kinase (SK)) lacking the Citron sequence. In keratinocytes, full-length CRIK, but not CRIK-SK, localizes into corpuscular cytoplasmic structures and recruits actin into these structures. Like the ubiquitously expressed Rho-associated kinases ROCK I and II kinases, Crik may participate in the dynamics of the actin cytoskeleton, albeit in a more tissue-restricted manner.

SGK064 (SEQ ID NO:59) is a new GRK family member that lists as its prototype the G protein-coupled receptor kinases (i.e. beta ARK1). Domain analysis of SGK064 revealed the presence of an RGS domain. The RGS (regulator of G protein signaling) domain (PF00615) is approximately 125 amino acids long and found in family members that include the GTPase-

activating proteins for heterotrimeric G-protein alpha-subunits. G protein-coupled receptor kinases function in the beta-adrenergic signaling cascade, an important regulator of myocardial function. Several cardiovascular diseases such as congestive heart failure display alterations in this pathway (Proc Assoc Am Physicians 1999 Sep-Oct;111(5):399-405). The predicted open  
5 reading frame for SGK064 is 84% identical to GRK7 isolated from *Spermophilus tridecemlineatus* (ground squirrel). From this observation, SGK064 may represent the human orthologue or a very close homolog of *Spermophilus tridecemlineatus* GRK7. GRK7 may play an important role in signal transduction in the cones and rods of the visual system. Desensitization in the rod cell of the mammalian retina is initiated when light-activated  
10 rhodopsin is phosphorylated by the G protein-coupled receptor kinase (GRK) or GRK1. Like GRK1, GRK7 may function as an opsin kinase in the mammalian signal phototransduction process.

SGK409 (SEQ ID NO:60) is a new MAST family member that list as its prototype the microtubule-associated testis specific serine/threonine protein (MAST205 ). Domain analysis of  
15 SGK409 revealed the presence of a PDZ domain. The PDZ domain (of approximately 83 amino acids in length) is found in membrane-associated proteins that include homologues of the MAGUK family of guanylate kinases, several protein phosphatases and protein kinases. PDZ domains are also found in neuronal nitric oxide synthase as well as in the subfamily of dystrophin-associated proteins, collectively known as syntrophins. SGK409 is 60% identical  
20 (over 1351 amino acids) to murine MAST205, a protein implicated in mammalian spermiogenesis (Biol Reprod. 1996 Nov;55(5):1039-44). MAST205 and the related SAST (syntrophin-associated serine/threonine kinase) associate with beta 2-syntrophin via PDZ-PDZ domain interactions. MAST205 colocalizes with beta 2-syntrophin and utrophin at neuromuscular junctions. (Nat Neurosci 1999 Jul;2(7):611-7). Like MAST205 and SAST,  
25 SGK409 may be an essential player at the neuromuscular junction by linking the dystrophin/utrophin network to microtubules via the syntrophins.

SGK021 (SEQ ID NO:61) is a new member Mo3C11.1\_ce family member named after the *c. elegans* Mo3C11 hypothetical open reading frame. One Mo3C11.1 family member, Tank-binding kinase (TBK1), has been studied with respect to its signaling properties. TBK1 is an  
30 IKK-related kinase capable of activating the NfκB pathway following its interaction with the

adaptor molecules TRAF2 and TANK (EMBO J 1999 Dec 1;18(23):6694-704). SGK021 may share signaling properties with TBK1.

SGK410 (SEQ ID NO:62) is a new PKC family member closest to PKC $\epsilon$  (95% identity over 587 amino acids). PKC $\epsilon$  has been implicated to function in the cancer cell survival pathway since it protects human leukemia cells against drug-induced apoptosis (J Biol Chem 1997 Oct 31;272(44):27521-4, J Biol Chem 1999 Feb 12;274(7):3927-30). Given the high degree of homology between SGK410 and PKC $\epsilon$ , it is conceivable that SGK410 functions in the cell survival pathway in cancer.

SGK069 (SEQ ID NO:63) and SGK110 (SEQ ID NO:64) are new AGC group members that do not cluster with other AGC kinases, hence are referred to as unique family members. The low degree of homology of SGK069 and SGK110 to their closest kinases (42 and 40%) precludes making any homology-based inferences on the potential signaling properties of these novel AGC group kinases. The 48 tissue PCR-based expression pattern for SGK069 showed it to be restricted both in normal (confined to skeletal muscle) and tumor tissues (kidney carcinoma). The same PCR-based expression panel for SGK110 showed this gene to be expressed in placenta, spleen, thyroid gland, uterus and lung. SGK110 was also expressed in multiple cancer cell lines (*see, e.g.*, Table 7).

### CAMK Group

Potential biological and clinical implications of the novel CAMK group protein kinases are described next. SGK053 (SEQ ID NO:65) and SGK124 (SEQ ID NO:66) are novel AMPK family members of the CAMK group. Since the initial filing of this application, the full-length version of human SGK053 appeared in the public database as CamKI-like protein kinase (CaMKI) (NP\_065130.1). CaMKI is a calcium-calmodulin-dependent kinase expressed in granulocytes whose transcription is upregulated during neutrophilic differentiation of CD34(+) stem cells. Activation of CaMKI induces extracellular signal-related kinase (ERK) mitogen-activated protein (MAP) kinase activity and CRE-binding protein (CREB) transcriptional activity (Blood. 2000;96:3215-3223). Since the initial filing of this application, the full-length version of human SGK124 appeared in the public database (CAB81634.1) as a hypothetical protein of unknown function.



SGK254 (SEQ ID NO:67), SGK297 (SEQ ID NO:68) and SGK411 (SEQ ID NO:69) are novel CAMK family members of the CAMK group. SGK254 is most likely the human orthologue or a very close homolog of rat Ca<sup>2+</sup>/calmodulin-dependent PK IV (CaM-kinase IV) (91% over 505 amino acids). CaM-kinase IV mediates Ca<sup>2+</sup>-dependent regulation through the phosphorylation of specific transcription factors in the central nervous and immune systems. CaM-kinase I and IV but not CaM-kinase II are activated by an upstream CaM-kinase kinase (J Biol Chem 1995 Aug 18;270(33):19320-4). Since the initial filing of this application, the full-length versions of human SGK297 and SGK 411 appeared in the public database as Ca<sup>2+</sup>/Calmodulin-dependent protein kinase I (CamKI)-like protein kinase (CaMKI (AAF74509.1) and multifunctional CAMK II delta2 (AAD20442.1), respectively. SGK297 was identified as a hypothetical ORF from genomic sequence from Xq28 (Genomics 1997 Aug 15;44(1):8-14). The putative rat orthologue of human SGK297, rat Ca<sup>2+</sup>/calmodulin (CaM)-dependent protein kinase I exists in at least three splicing isoforms (CaM kinase, Ibeta1 and Igamma) that provide alternative means of Ca<sup>2+</sup>/Calmodulin-dependent signaling in the nervous system (J Biol Chem 1997 Dec 19;272(51):32704-8). SGK 411 encodes multifunctional CAMK II delta2, a kinase potentially implicated in heart failure (Circ Res 1999 Apr 2;84(6):713-21).

SGK004 (SEQ ID NO:75), SGK006 (SEQ ID NO:76) and SGK180(SEQ ID NO:77) are novel CAMK group members belonging to the SNF1 subfamily of EMK-1 kinases. SGK386 (SEQ ID NO:78) is a novel CAMK group member belonging to the MLCK family.

Since the initial filing of this application, SGK004 and SGK180 appeared in the public database as the full-length gene encoding SNF1LK (P57059) and HSNFRK (AAF86944.1), respectively. SGK006 is 46% identical over 121 amino acids to mouse hormonally upregulated Neu-associated kinase (Hunk or MAK-V) (NP\_056570.1), a hypothetical protein predicted from the entire sequence of human chromosome 21 (Nature 405 (6784), 311-319 (2000)). SGK004, SGK006 and SGK180 display homology to human Hunk (NP056570). These three kinases also display significant homology to rat salt-induced kinase (SIK, also known as KID2 protein) (NP\_067725.1), mouse myocardial SNF1-like kinase (MSK)( NP\_034961.1) and chicken Qik (qin-induced kinase)( AAF28351). From yeast to plants and mammals, the SNF1 protein kinase plays a central role in stress-response pathways (Annu Rev Biochem 1998;67:821-55). The ubiquitously expressed Qik protein may play a role in oncogenesis (Biochem Biophys Res

Commun 2000 Sep 24;276(2):564-70Z) and Hunk (MAK-V) may have a role in pregnancy and mammary carcinogenesis (Development 2000 Oct;127(20):4493-509). SGK004, SGK006 and SGK180 may fulfill roles in the response to cellular stress. Disruptions in the signaling pathways in which any of these novel kinases participate may trigger cancer or other disease conditions.

5

SGK386 (SEQ ID NO:78) is another CAMK group member presented in this application. Since the initial filing of this application, SGK386 appeared in the public database as the full-length gene encoding myosin, light polypeptide kinase (MLCK) (CAC10006.1). MLCK is responsible for regulating the contractility of actomyosin fibers, both in smooth muscle as well as in non-muscle vertebrate cells. MLCK has been implicated in cell growth regulation through its ability to modulate myosin II motor activity via protein phosphorylation, thereby triggering cytokinesis (J Cell Biol 2000 Oct 30;151(3):697-708). Disruptions in SGK386 signaling may also be involved in triggering cancer or other disease conditions.

10

#### 15 **Casein kinase group**

SGK003 (SEQ ID NO:79) and SGK006 (SEQ ID NO:80) are novel casein kinases that belong to the CKI family. SGK003 is 91% identical over 304 amino acids to casein kinase I, alpha 1. The high degree of sequence homology between SGK003 and CKI, alpha suggests that these two kinases may have overlapping functions in processes related to vesicular trafficking, DNA repair, cell cycle progression or cytokinesis. Noteworthy is the elevated expression levels observed for SGK003 in tumor versus normal tissue sources as determined from both, the tissue array and PCR-based expression studies. SGK003 was found to be expressed in the following normal sources: spinal cord, uterus, fetal brain, fetal kidney, fetal lung and adrenal gland. SGK003 was overexpressed in tumors associated with the following tissues: colon, lung, prostate, ovary, bone, breast and brain (glioblastoma). SGK003, therefore, may be a potential target for therapeutical intervention in oncology. SGK006, the second novel CKI family member described in this application is 53% identical over 166 amino acids to the hypothetical open reading frame R90.1 from c. elegans.

20

25

#### 30 **CMGC kinase group**

SGK041 (SEQ ID NO:81) and SGK112 (SEQ ID NO:82) are novel CMGC protein kinases that belong to the CDK family. Since the initial filing of this application, SGK041 appeared in the public database as the full-length gene encoding Nkiamre (NP\_057592.1). Nkiamre is the human orthologue or a very close homolog of the rat protein Nkiatre. Nkiamre was found to be deleted at both alleles in 9 of 18 leukemic samples with chromosome band 5q31 abnormalities revealed by fluorescence in situ chromosomal hybridization (FISH) (Cancer Res 1999 Aug 15;59(16):4069-74). SGK041 may have tumor suppressor activities required for normal hematopoiesis. Loss of SGK041 and/or deregulation of its signalling pathway may be implicated in the etiology of human acute leukemia and myelodysplasia (Cancer Res 1999 Aug 15;59(16):4069-74). SGK112, the second novel CDK family member in this application is 61% related to the human CDC2-related kinase (CCRK) whose function is unknown. On the basis of sequence similarity to cyclin-dependent kinase 2 (Cdk2 (44% amino acid identity over 480 amino acids), SGK112 may have cell cycle regulatory functions akin to those played by Cdk2. Cdk2 has been implicated in the etiology of non-small cell lung carcinoma (Cell Growth Differ 2000 Oct;11(10):507-15). Disruptions in SGK112 signaling may also be involved in triggering cancer or other disease conditions.

SGK038 (SEQ ID NO:83), is most likely the human orthologue or a very close homolog of rat ERK7 (P42525). ERK7 is a nuclear-localized, constitutively active MAP kinase capable of inhibiting cellular growth. In spite of the presence of the canonical TEY catalytic motif found in other MAP kinases, ERK7 is not responsive to the mitogenic stimuli that normally activate the MAP kinase pathways. The kinase activity, growth-inhibitory and nuclear localization properties of ERK7 are dependent on an intact C-terminal extracatalytic region (Mol Cell Biol 1999 Feb;19(2):1301-12). The C-terminus of ERK7 has been shown to bind through its C-terminal region to an intracellular chloride ion (CLIC3). (J Biol Chem 1999 Jan 15;274(3):1621-7). Disruptions in SGK038 signaling may also be involved in triggering cancer or other disease conditions.

#### **Microbial PK group**

SGK158 (SEQ ID NO:84) and SGK429 (SEQ ID NO:85) are new members of the ATP binding cassette transporter (ABC) family of mammalian microbial-like kinases. Since the initial

filing of this application, SGK158 and SGK429 appeared in the public database as the full-length gene encoding two hypothetical proteins predicted from genomic sequence (NP\_064632 and AAD43192.1, respectively). One mammalian ABC protein, ABC1 has been studied with respect to its physiological role. ABC1 is required for cell engulfment by macrophages undergoing  
5 apoptosis. Other members of the ABC1 family of mammalian proteins include the cystic fibrosis transmembrane conductance (CFTR) regulator and as P-glycoprotein. Like CFTR and P-glycoprotein, ABC1 is an anion transporter. In microbes, ABC transporter play an important role in multidrug resistance (Biochim Biophys Acta 1998 Jun 10;1365(1-2):31-6). Like other ABC protein, SGK158 and SGK429 may have a function in the mechanism of drug resistance.  
10 SGK152 (SEQ ID NO:86), the third novel microbial-like kinase, belongs to the R101 family. Since the initial filing of this application, SGK152 appeared in the public database as the full-length gene encoding the human homolog of *Aspergillus nidulans* sudD (suppressor of bimD6) (NP\_003822.1). sudD is conserved from yeast to mammals and appears to play a key function in cell cycle regulation related to chromosome condensation and segregation (Gene  
15 1998 May 12;211(2):323-9). Disruptions in SGK152 signaling may also be involved in triggering cancer or other disease conditions.

#### **“Other” group**

SGK077 (SEQ ID NO:87), a novel C26C2 family member, is closest to murine hapsin  
20 (BAB00640.1) (62% amino acid identity over 492 amino acids). Both SGK077 and hapsin have motifs predicted from a hidden Markov model analysis that suggests that these proteins act as kinases. The biological function of hapsin is unknown.

SGK093 (Wnk3) (SEQ ID NO:88), a novel C26C2 family member, is closest to kinase-deficient protein (KDP) (AAF31483.1) (60% amino acid identity over 321 amino acids). Wnk3  
25 is a member of a subfamily of serine/threonine kinases which includes a described prototype, Wnk1, isolated from rat (J Biol Chem 2000 Jun 2;275(22):16795-801). This family is characterized by an N-terminal catalytic domain with several unique sequence features, most notably a change of the invariant lysine in kinase subdomain II to a cysteine, coupled with a change of the third conserved glycine residue in subdomain I into a lysine. The resulting enzyme  
30 appears to maintain catalytic activity due to this concomitant switch. Wnk3 conserves both of

these catalytic changes and therefore is predicted to maintain catalytic activity. The long C-terminal portion of the Wnks includes many protein interaction domains such as SH3 binding sites and coiled coil regions.

The Wnk family catalytic domain shows the highest similarity to two families of serine/threonine kinases: The MEKK-like kinases and the Ste20-like kinases. Both of these families can regulate enzymes in various MAPK signaling cascades, which are critical for many cellular processes such as mitogenesis, differentiation, cell survival, and stress response. The Ste20 kinases are also involved in regulation of the ras/rac/rho/cdc42 pathways and subsequent downstream effects on cytoskeleton. Wnk3 shows high expression in human kidney, in kidney carcinoma cell lines, in prostate, prostate cell lines, and prostate tumor bone metastases, in colorectal tissue and tumor cell lines, and in human leukemia cells. Therefore Wnk3 may be involved in the normal homeostasis and functioning of the human kidney, prostate, and digestive system, and may be involved in tumorigenesis which arises from these three tissues. High expression in human leukemia cell lines indicates a possible role in the development of that disease as well.

The partial SGK074 (SEQ ID NO:89), a novel Dyrk family member, is closest to the murine Myak protein kinase (AAD41593.1) with 50% amino acid sequence identity over 176 amino acids. Murine Myak is the homolog of the yeast kinase Yak1. The Yak1 family of related genes includes drosophila (Mnb) and human minibrain (Dyrk1a) and the rat steroid hormone receptor interacting protein ANPK. Myak features a nuclear localization signal and PEST motifs flanking its catalytic domain (Mol Reprod Dev 2000 Apr;55(4):372-8). SGK074 may represent an additional member of steroid-responsive protein kinases.

SGK087 (SEQ ID NO:90) and SGK295, Kis, P-CIP2 (SEQ ID NO:91) represent the two additional Dyrk family members in this application. Since the initial filing of this application, SGK087 and SGK295 appeared in the public database as the full-length gene Dyrk4 (AAF91393.1) (predicted from genomic sequence) and the rat Kis protein kinase (Kinase interacting with stathmin (P-CIP2) (NP\_058989.1). Dyrk is the fourth member of the human Dyrk family of protein kinases, the other three being Dyrk1 (human minibrain), Dyrk2 and Dyrk3. The functional role of Dyrk's is presently unknown. P-CIP2 was found to be an interacting protein for the cytosolic COOH-terminal domain (CD) of peptidylglycine alpha-

amidating monooxygenase (PAM), a key element in the secretory pathway of neurons and endocrine cells (J Biol Chem 1996 Nov 8;271(45):28636-40).

SGK419 (SEQ ID NO:92) represents a novel NAK family member within the Other group. Since the initial filing of this application, SGK419 appeared in the public database as the full-length gene encoding the hypothetical protein CAB70863.1 predicted from human genomic sequencing. SGK419 is closely related to drosophila Numb-associated kinase (NAK) (65% sequence identity over 294 amino acids). In drosophila, NAK, through its interaction with the membrane-associated protein Numb, participates in cell fate determination during asymmetric cell divisions (Mol Cell Biol 1998 Jan;18(1):598-607). SGK419 may play a similar role as NAK in processes related to cell differentiation and proliferation.

SGK125, Myo3A (SEQ ID NO:93) represents a novel NinaC family kinase within the Other group. Since the initial filing of this application, SGK125 appeared in the public database as the full-length gene encoding MyoIIIA (NP\_059129). SGK125 displays 34% sequence identity over 1155 amino acids to drosophila Ninac, another class III myosin. Like drosophila NinaC, Myo3A displays a myosin head domain as well as 3 calmodulin-binding (IQ) domains. Ninac plays a key role in the photoreceptor light signal transduction process (Biochim Biophys Acta 2000 Mar 17;1496(1):52-9. SGK125 may represent a functional counterpart of Ninac with roles in mammalian phototransduction.

The partial SGK445 (SEQ ID NO:94) represents a novel polo kinase (PLK) family kinase within the Other group. SGK445 is closest to murine Sak (AAC37649.1), a drosophila polo-like kinase, with 65% sequence identity over 47 amino acids. The close homology of SGK445 to Sak suggests that SGK445 may represent the human orthologue of the murine kinase. Sak expression is associated with the processes of mitosis and meiosis. The fruit fly and mammalian forms of PLK have been shown to play a key role in cell cycle progression (Curr Opin Cell Biol 1998 Dec;10(6):776-83). SGK445, like Sak and PLK, may play similar roles during cell proliferation.

SGK127 (SEQ ID#95) represents a novel RAF family kinase within the Other group.

SGK127 is closest to KSR (kinase suppressor of Ras) NP\_038599.1 with 47% sequence identity over 446 amino acids. KSR has been shown to function in the Ras-mediated signal transduction of multiple receptor tyrosine kinase- activated pathways. Hence, SGK127 may be

implicated in cell growth regulation. Disruptions in SGK127 signaling may be involved in triggering cancer or other disease conditions

SGK009, Ankrd3 (SEQ ID NO:96) represents a novel RIP family kinase within the Other group. Since the initial filing of this application, SGK009 appeared in the public database as the full-length gene encoding Ankrd3 (NP\_065690.1) predicted from human genomic sequence from chromosome 21 (Nature 2000 May 18;405(6784):311). SGK009 features 10 extracatalytic C-terminal ankyrin domains. The biological function of Ankrd3 is unknown. The presence of multiple ankyrin domains in SGK009 (Ankrd3) suggests that this protein plays an important scaffolding role akin to that observed in the integrin-like kinases (Int J Mol Med 1999 Jun;3(6):563-72). Such scaffolding kinases participate in integrin-, growth factor- and Wnt-signaling pathways that are important in normal as well as tumor cell proliferation.

SGK421, STK22A, TSK1 (SEQ ID NO:97) represents a novel STK22A family kinase within the Other group. Since the initial filing of this application, the nucleic sequence corresponding to SGK421 appeared in the public database from genomic sequence (AC008476). Based on sequence homology, SGK421 appears to be the human orthologue of the murine kinase STK22A (NP\_033461.1). TSK1 maps to the DiGeorge critical region (DGCR) human chromosome 22q11 and is syntenic with the STK22A gene in the mouse genome (chromosome 16). Patients with DiGeorge syndrome have velocardiofacial or conotruncal facial anomalies (J Otolaryngol 2000 Sep-Oct;21(5):326-30). SGK421 may be implicated in the etiology of DiGeorge syndrome.

The partial SGK047 (SEQ ID NO:98) represents a novel STKR family kinase within the Other group. SGK047 is closest to bone morphogenetic protein (BMP) receptor, type IA (NP\_004320.1) with 46% sequence identity over 11 amino acids. The full-length form of SGK047 may potentially encode a novel bone morphogenetic protein receptor. BMP receptors play a key role in embryogenesis (Cell Mol Life Sci. 2000 Jun;57(6):943-56.).

SGK196 (SEQ ID NO:99) represents a novel unique family kinase within the Other group. Since the initial filing of this application, SGK196 appeared in the public database as the full-length gene encoding the hypothetical ORF BAB15623.1. SGK196 is closest to multiple protein kinases from Arabidopsis thaliana sharing 28% sequence identity over 153 amino acids. The function of the plant kinases is unknown.

SGK396 (SEQ ID NO:100) represents a novel unique family kinase within the Other group. SGK396 is closest to the *Arabidopsis thaliana* kinase (BAB11570.1) with 28% amino acid sequence identity over 34 amino acids. The function of the plant kinase is unknown.

5        SGK279, PKN (SEQ ID NO:101) represents a novel YWY3<sub>ce</sub> family kinase within the Other group. SGK279 is closest to the rat kinase encoded by BAA36362.1 with 94% sequence identity over 400 amino acids. The function of the rat kinase is unknown.

#### The STE Group

10        SGK037 (SEQ ID NO:102), SGK060 (SEQ ID NO:103) and SGK080 (SEQ ID NO:104) represent three novel NEK family members of the STE group. SGK037 is closest to NEK1 with 65% sequence identity over 176 amino acids. Since the initial filing of this application, SGK060 appeared in the public database as the full-length gene encoding the hypothetical protein hypothetical BAB15672.1. SGK080 is closest to NRK (NP\_002488.1) with 90% sequence  
15        identity over 404 amino acids. NEK family kinases such as NEK1 and NRK are related to the mitotic regulator NimA from *Aspergillus nidulans*. Given the homology of SGK037, SGK060 and SGK080 to various NEK family members, these novel kinases may participate in cell cycle regulation. Disruptions in the signaling pathways in which SGK037, SGK060 and SGK080 participate may be associated with cancer or other diseases.

20        SGK002 (SEQ ID NO:105), SGK058 (SEQ ID NO:106) and SGK103 (SEQ ID NO:107) represent three novel STE11 family members of the STE group. SGK002 (SEQ ID NO:105) is closest to MEK2 (P36507) with 92% sequence identity over 369 amino acids. A second MEK-like kinase is represented by SGK058. Since the initial filing of this application, SGK058 appeared in the public database as the full-length gene encoding the hypothetical protein  
25        BAB15538.1. Given the high degree of similarity between SGK002 and MEK2, SGK002 may be implicated in MAP kinase pathway regulation.

The partial SGK103 (SEQ ID NO:107) represents the third STE11 family kinase of this application. SGK103 is closest to *drosophila* fused (CAA39285.1) with 57% sequence homology over 16 amino acids.



SGK035 (SEQ ID NO:108) and SGK075 (SEQ ID NO:109) represent two novel STE20 family members of the STE group. SGK035 display 92% over 318 amino acids to PAK2 Q13177. SGK075 displays 67% over 227 amino acids to myosin IIIA. The STE20 family of protein kinases represent key membrane-proximal regulators of multiple signal transduction pathways important in cell proliferation, survival, differentiation and response to cellular stress, all potential functions for the polypeptides represented by SGK035 and SGK075.

#### The TK group

SGK188 (SEQ ID NO:110) is a novel receptor tyrosine kinase belonging to the EphA9 family. The Eph family is the largest subfamily of receptor tyrosine kinases in the human genome. This family has a stereotyped structure consisting of an N-terminal globular domain involved in ligand binding, two Type III fibronectin-like domains which contribute to receptor dimerization, a transmembrane domain, and an intracellular tyrosine kinase domain.

SGK040 (SEQ ID NO:54), the second novel member of the tyrosine kinase group, belongs to the unique family since it fails to cluster with other tyrosine kinase family members. SGK040 displays 30% sequence identity over 76 amino acids to the rat cytoplasmic tyrosine kinase (CTK) BATK (P41243). Given the pivotal role of CTK's in signal transduction (Int J Mol Med. 2000 Jun;5(6):583-90, Curr Pharm Des. 2000 Mar;6(4):361-78), disruptions in SGK040 signaling may be involved in triggering cancer or other disease conditions.

20

#### **EXAMPLE 2f: Classification of polypeptides exhibiting kinase-like activity among defined groups**

#### **DAG Kinase Group**

25

SGK390 (SEQ ID NO:112) represents a novel family member of the DAG family of kinases. Since the initial filing of this application, a potential orthologue or very close homolog from *Mesocricetus auratus* (hamster) of human SGK390 appeared in the public database as the full-length gene encoding diacylglycerol kinase eta (DAG kinase eta) (NP\_057592.1). Like its hamster counterpart, SGK390 contains multiple extracatalytic domains defined from a profile

30

analysis, including a diacylglycerol/phorbol ester binding domain, a plekstrin domain and a diacylglycerol kinase accessory domain. Hamster DAG kinase eta was shown to have a wide tissue distribution in contrast to other DAG kinases which display a more restricted tissue distribution pattern. DAG kinases have been shown to play a key role in regulating the concentration of the second messenger DAG (J Biol Chem 1996 Aug 16;271(33):19781-8). Given the potential role of SGK390 in regulating DAG levels, disruptions in the signaling pathway in which this kinase participates may trigger cancer or other disease conditions.

### Guanylate Cyclases (GCyc) Group

SGK007 (SEQ ID NO:113) and SGK050 (SEQ ID NO:114) are novel members of the guanylate cyclase family of protein kinase-like molecules. SGK007 is 59% identical over 499 amino acids to rat guanylate cyclase; hence, SGK007 is the human homologue or a close homolog of the rat protein. The partial SGK050 polypeptide is 48% identical over 23 amino acids to the natriuretic peptide receptor B precursor, isoform b (NP\_003986.2). A similar homology is found between SGK050 and guanylate cyclase from multiple species.

### EXAMPLE 3: Isolation of cDNAs Encoding Mammalian Protein Kinases

#### Materials and Methods

##### Identification of novel clones

Total RNAs are isolated using the Guanidine Salts/Phenol extraction protocol of Chomczynski and Sacchi (P. Chomczynski and N. Sacchi, *Anal. Biochem.* 162, 156 (1987)) from primary human tumors, normal and tumor cell lines, normal human tissues, and sorted human hematopoietic cells. These RNAs are used to generate single-stranded cDNA using the Superscript Preamplification System (GIBCO BRL, Gaithersburg, MD; Gerard, GF *et al.* (1989), *FOCUS* 11, 66) under conditions recommended by the manufacturer. A typical reaction uses 10 µg total RNA with 1.5 µg oligo(dT)<sub>12-18</sub> in a reaction volume of 60 µL. The product is treated with RNaseH and diluted to 100 µL with H<sub>2</sub>O. For subsequent PCR amplification, 1-4 µL of this sscDNA is used in each reaction.

Degenerate oligonucleotides are synthesized on an Applied Biosystems 3948 DNA synthesizer using established phosphoramidite chemistry, precipitated with ethanol and used unpurified for PCR. These primers are derived from the sense and antisense strands of conserved motifs within the catalytic domain of several protein kinases. Degenerate nucleotide residue designations are: N = A, C, G, or T; R = A or G; Y = C or T; H = A, C or T not G; D = A, G or T not C; S = C or G; and W = A or T.

PCR reactions are performed using degenerate primers applied to multiple single-stranded cDNAs. The primers are added at a final concentration of 5  $\mu$ M each to a mixture containing 10 mM TrisHCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 200  $\mu$ M each deoxynucleoside triphosphate, 0.001% gelatin, 1.5 U AmpliTaq DNA Polymerase (Perkin-Elmer/Cetus), and 1-4  $\mu$ L cDNA. Following 3 min denaturation at 95 °C, the cycling conditions are 94 °C for 30 s, 50 °C for 1 min, and 72 °C for 1 min 45 s for 35 cycles. PCR fragments migrating between 300-350 bp are isolated from 2% agarose gels using the GeneClean Kit (Bio101), and T-A cloned into the pCRII vector (Invitrogen Corp. U.S.A.) according to the manufacturer's protocol.

Colonies are selected for mini plasmid DNA-preparations using Qiagen columns and the plasmid DNA is sequenced using a cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products are run on an ABI Prism 377 DNA Sequencer, and analyzed using the BLAST alignment algorithm (Altschul, S.F. *et al.*, *J.Mol.Biol.* 215: 403-10).

Additional PCR strategies are employed to connect various PCR fragments or ESTs using exact or near exact oligonucleotide primers. PCR conditions are as described above except the annealing temperatures are calculated for each oligo pair using the formula:  $T_m = 4(G+C)+2(A+T)$ .

#### Isolation of cDNA clones:

Human cDNA libraries are probed with PCR or EST fragments corresponding to kinase-related genes. Probes are <sup>32</sup>P-labeled by random priming and used at 2x10<sup>6</sup> cpm/mL following standard techniques for library screening. Pre-hybridization (3 h) and hybridization (overnight) are conducted at 42 °C in 5X SSC, 5X Denhart's solution, 2.5% dextran sulfate, 50 mM Na<sub>2</sub>PO<sub>4</sub>/NaHPO<sub>4</sub>, pH 7.0, 50% formamide with 100 mg/mL denatured salmon sperm DNA.

Stringent washes are performed at 65 °C in 0.1X SSC and 0.1% SDS. DNA sequencing was carried out on both strands using a cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products are run on an ABI Prism 377 DNA Sequencer.

**EXAMPLE 4: Expression Analysis of Mammalian Protein Kinases****Materials and Methods****Northern blot analysis**

5 Northern blots are prepared by running 10 µg total RNA isolated from 60 human tumor cell lines (such as HOP-92, EKVX, NCI-H23, NCI-H226, NCI-H322M, NCI-H460, NCI-H522, A549, HOP-62, OVCAR-3, OVCAR-4, OVCAR-5, OVCAR-8, IGROV1, SK-OV-3, SNB-19, SNB-75, U251, SF-268, SF-295, SF-539, CCRF-CEM, K-562, MOLT-4, HL-60, RPMI 8226, SR, DU-145, PC-3, HT-29, HCC-2998, HCT-116, SW620, Colo 205, HTC15, KM-12, UO-31, 10 SN12C, A498, CaKi1, RXF-393, ACHN, 786-0, TK-10, LOX IMVI, Malme-3M, SK-MEL-2, SK-MEL-5, SK-MEL-28, UACC-62, UACC-257, M14, MCF-7, MCF-7/ADR RES, Hs578T, MDA-MB-231, MDA-MB-435, MDA-N, BT-549, T47D), from human adult tissues (such as thymus, lung, duodenum, colon, testis, brain, cerebellum, cortex, salivary gland, liver, pancreas, kidney, spleen, stomach, uterus, prostate, skeletal muscle, placenta, mammary gland, bladder, 15 lymph node, adipose tissue), and 2 human fetal normal tissues (fetal liver, fetal brain ), on a denaturing formaldehyde 1.2% agarose gel and transferring to nylon membranes.

Filters are hybridized with random primed [ $\alpha^{32}\text{P}$ ]dCTP-labeled probes synthesized from the inserts of several of the kinase genes. Hybridization is performed at 42 °C overnight in 6X SSC, 0.1% SDS, 1X Denhardt's solution, 100 µg/mL denatured herring sperm DNA with 1-2 x 20  $10^6$  cpm/mL of  $^{32}\text{P}$ -labeled DNA probes. The filters are washed in 0.1X SSC/0.1% SDS, 65 °C, and exposed on a Molecular Dynamics phosphorimager.

**Quantitative PCR analysis**

RNA is isolated from a variety of normal human tissues and cell lines. Single stranded 25 cDNA is synthesized from 10 µg of each RNA as described above using the Superscript Preamplification System (GibcoBRL). These single strand templates are then used in a 25 cycle PCR reaction with primers specific to each clone. Reaction products are electrophoresed on 2% agarose gels, stained with ethidium bromide and photographed on a UV light box. The relative intensity of the STK-specific bands were estimated for each sample.

30

**DNA Array Based Expression Analysis**

Plasmid DNA array blots are prepared by loading 0.5 µg denatured plasmid for each kinase on a nylon membrane. The [ $\gamma^{32}\text{P}$ ]dCTP labeled single stranded DNA probes are synthesized from the total RNA isolated from several human immune tissue sources or tumor cells (such as thymus, dendrocytes, mast cells, monocytes, B cells (primary, Jurkat, RPMI8226, SR), T cells (CD8/CD4+, TH1, TH2, CEM, MOLT4), K562 (megakaryocytes). Hybridization is performed at 42 °C for 16 hours in 6X SSC, 0.1% SDS, 1X Denhardt's solution, 100 µg/mL denatured herring sperm DNA with 10<sup>6</sup> cpm/mL of [ $\gamma^{32}\text{P}$ ]dCTP labeled single stranded probe. The filters are washed in 0.1X SSC/0.1% SDS, 65 °C, and exposed for quantitative analysis on a Molecular Dynamics phosphorimager.

10

#### **EXAMPLE 5: Protein Kinase Gene Expression**

##### Vector Construction

##### Materials and Methods

15

##### Expression Vector Construction

Expression constructs are generated for some of the human cDNAs including: a) full-length clones in a pCDNA expression vector; b) a GST-fusion construct containing the catalytic domain of the novel kinase fused to the C-terminal end of a GST expression cassette; and c) a full-length clone containing a Lys to Ala (K to A) mutation at the predicted ATP binding site within the kinase domain, inserted in the pCDNA vector.

20

The "K to A" mutants of the kinase might function as dominant negative constructs, and will be used to elucidate the function of these novel STKs.

#### **EXAMPLE 6: Generation of Specific Immunoreagents to Protein Kinases**

25

##### Materials and Methods

Specific immunoreagents are raised in rabbits against KLH- or MAP-conjugated synthetic peptides corresponding to isolated kinase polypeptides. C-terminal peptides were conjugated to KLH with glutaraldehyde, leaving a free C-terminus. Internal peptides were MAP-conjugated with a blocked N-terminus. Additional immunoreagents can also be generated by immunizing rabbits with the bacterially expressed GST-fusion proteins containing the cytoplasmic domains of each novel PTK or STK.

30

The various immune sera are first tested for reactivity and selectivity to recombinant protein, prior to testing for endogenous sources.

#### Western blots

5 Proteins in SDS PAGE are transferred to immobilon membrane. The washing buffer is PBST (standard phosphate-buffered saline pH 7.4 + 0.1% Triton X-100). Blocking and antibody incubation buffer is PBST +5% milk. Antibody dilutions varied from 1:1000 to 1:2000.

### EXAMPLE 7: Recombinant Expression and Biological Assays for Protein Kinases

10

#### Materials and Methods

#### Transient Expression of Kinases in Mammalian Cells

The pcDNA expression plasmids (10 µg DNA/100 mm plate) containing the kinase constructs are introduced into 293 cells with lipofectamine (Gibco BRL). After 72 hours, the  
15 cells are harvested in 0.5 mL solubilization buffer (20 mM HEPES, pH 7.35, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1.5 mM MgCl<sub>2</sub>, 1 mM EGTA, 2 mM phenylmethylsulfonyl fluoride, 1 µg/mL aprotinin). Sample aliquots are resolved by SDS polyacrylamide gel electrophoresis (PAGE) on 6% acrylamide/0.5% bis-acrylamide gels and electrophoretically transferred to nitrocellulose. Non-specific binding is blocked by preincubating blots in Blotto  
20 (phosphate buffered saline containing 5% w/v non-fat dried milk and 0.2% v/v nonidet P-40 (Sigma)), and recombinant protein was detected using the various anti-peptide or anti-GST-fusion specific antisera.

#### In Vitro Kinase Assays

25 Three days after transfection with the kinase expression constructs, a 10 cm plate of 293 cells is washed with PBS and solubilized on ice with 2 mL PBSTDS containing phosphatase inhibitors (10 mM NaHPO<sub>4</sub>, pH 7.25, 150 mM NaCl, 1% Triton X-100, 0.5% deoxycholate, 0.1% SDS, 0.2% sodium azide, 1 mM NaF, 1 mM EGTA, 4 mM sodium orthovanadate, 1% aprotinin, 5 µg/mL leupeptin). Cell debris was removed by centrifugation (12000 x g, 15 min, 4  
30 °C) and the lysate was precleared by two successive incubations with 50 µL of a 1:1 slurry of protein A sepharose for 1 hour each. One-half mL of the cleared supernatant was reacted with

10  $\mu$ L of protein A purified kinase-specific antisera (generated from the GST fusion protein or anti-peptide antisera) plus 50  $\mu$ L of a 1:1 slurry of protein A-sepharose for 2 hr at 4 °C. The beads were then washed 2 times in PBSTDS, and 2 times in HNTG (20 mM HEPES, pH 7.5/150 mM NaCl, 0.1% Triton X-100, 10% glycerol).

5       The immunopurified kinases on sepharose beads are resuspended in 20  $\mu$ L HNTG plus 30 mM  $MgCl_2$ , 10 mM  $MnCl_2$ , and 20  $\mu$ Ci [ $\alpha$ - $^{32}P$ ]ATP (3000 Ci/mmol). The kinase reactions are run for 30 min at room temperature, and stopped by addition of HNTG supplemented with 50 mM EDTA. The samples are washed 6 times in HNTG, boiled 5 min in SDS sample buffer and analyzed by 6% SDS-PAGE followed by autoradiography. Phosphoamino acid analysis is  
10 performed by standard 2D methods on  $^{32}P$ -labeled bands excised from the SDS-PAGE gel.

Similar assays are performed on bacterially expressed GST-fusion constructs of the kinases.

#### 15 **EXAMPLE 8a: Chromosomal Localization of Protein Kinases**

##### **Materials and Methods**

20 Several sources were used to find information about the chromosomal localization of each of the genes described in this patent. First, cytogenetic map locations of these contigs were found in the title or text of their Genbank record, or by inspection through the NCBI human genome map viewer ([http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/hum\\_srch?](http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/hum_srch?)). Alternatively, the accession number for the nucleic acid sequence was used to query the Unigene database. The site containing the Unigene search engine is: <http://www.ncbi.nlm.nih.gov/UniGene/Hs.Home.html>.  
25 Information on map position within the Unigene database is imported from several sources, including the Online Mendelian Inheritance in Man (OMIM, <http://www.ncbi.nlm.nih.gov/Omim/searchomim.html>), The Genome Database (<http://gdb.infobiogen.fr/gdb/simpleSearch.html>), and the Whitehead Institute human physical map ([http://carbon.wi.mit.edu:8000/cgi-bin/contig/sts\\_info?database=release](http://carbon.wi.mit.edu:8000/cgi-bin/contig/sts_info?database=release)).

30       Once a cytogenetic region has been identified by one of these approaches, disease association is established by searching OMIM with the cytogenetic location. OMIM maintains a



searchable catalog of cytogenetic map locations organized by disease. A thorough search of available literature for the cytogenetic region is also made using Medline (<http://www.ncbi.nlm.nih.gov/PubMed/medline.html>). References for association of the mapped sites with chromosomal abnormalities found in human cancer can be found in: Knuutila, et al.,  
 5 Am J Pathol, 1998, 152:1107-1123.

## Results

The chromosomal regions for mapped genes are listed below. The chromosomal  
 10 positions were cross-checked with the Online Mendelian Inheritance in Man database (OMIM, <http://www.ncbi.nlm.nih.gov/htbin-post/Omim>), which tracks genetic information for many human diseases, including cancer. References for association of the mapped sites with chromosomal abnormalities found in human cancer can be found in: Knuutila, et al., Am J Pathol, 1998, 152:1107-1123. A third source of information on mapped positions was searching  
 15 published literature (at NCBI, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) for documented association of the mapped position with human disease.

SGK187, CRIK (SEQ ID NO:1, ENCODING SEQ ID NO:58 maps to chromosomal position 12q24.23. Amplification of this chromosomal position has been associated with non-  
 20 small cell lung cancer (12q24.1-24.3; 2/50) (Knuutila, et al.).

SGK064, GRK7 (SEQ ID NO:2, ENCODING SEQ ID NO:59 maps to chromosomal position 3q24. This chromosomal position has been associated with squamous cell carcinomas of the head and neck (3/30) and Uterine cervix cancer (3/10). (Knuutila, et al.); Usher syndrome (OMIM, 276902 USHER SYNDROME, TYPE III; USH3)

25 SGK409, KIAA0303 (SEQ ID NO:3, ENCODING SEQ ID NO:60 maps to chromosomal position 5q12.1. This chromosomal position has been associated with cancer of the testis (15q15-qter; 2/11) (Knuutila, et al.).

SGK021 (SEQ ID NO:4, ENCODING SEQ ID NO:61 maps to chromosomal position 5q31.2. This chromosomal position has been associated with chondrosarcoma (2/45), (Knuutila,  
 30 et al.).

SGK410 (SEQ ID NO:5, ENCODING SEQ ID NO:62 maps to chromosomal position Xq23. This chromosomal position has been associated with cancer of the prostate (Xq23-qter; 1/9), (Knuutila, et al.), and X-linked mental retardation (OMIM, 300046 MENTAL RETARDATION, X-LINKED NONSPECIFIC, 23; MRX23).

5       SGK069 (SEQ ID NO:6, ENCODING SEQ ID NO:63 maps to chromosomal position 19p11-p13. This chromosomal position has been associated with small cell lung cancer (19p12, 2/22), (Knuutila, et al.).

SGK110 (SEQ ID NO:7, ENCODING SEQ ID NO:64 maps to chromosomal position 19q13.4. This chromosomal position has been associated with the following human  
10       diseases: cancer of the breast (19q13.1-qter; 1/33), (Knuutila, et al.).

SGK053, CKLiK (SEQ ID NO:8, ENCODING SEQ ID NO:65 maps to chromosomal position 10p14. This chromosomal position has been associated with familial arrhythmogenic right ventricular dysplasia (OMIM, 604401 ARRHYTHMOGENIC RIGHT VENTRICULAR DYSPLASIA, FAMILIAL).

15       SGK124 (SEQ ID NO:9, ENCODING SEQ ID NO:66 maps to chromosomal position 20p12.2-p13. This chromosomal position has been associated with the following human diseases: cancer of the gastroesophageal junction (20p12; 3/28), (Knuutila, et al.), familial noncompaction of left ventricle (OMIM, 604169 NONCOMPACTION OF LEFT VENTRICULAR MYOCARDIUM, FAMILIAL ISOLATED, AUTOSOMAL DOMINANT).

20       SGK254, CAMKKa (SEQ ID NO:10, ENCODING SEQ ID NO:67 maps to chromosomal position 17p13.3. This chromosomal position has been associated with the following human diseases: Lost in cervical cancer (loss of heterogeneity, [Lazo, The molecular genetics of cervical carcinoma. Br J Cancer. 1999 Aug;80(12):2008-18. Review].

25       SGK297, CaMKIb2 (SEQ ID NO:11, ENCODING SEQ ID NO:68 maps to chromosomal position Xq28. Translocations involving this chromosomal position has been associated with the following human diseases: T cell prolymphocytic leukemia (Laine, et al.Mol Cell. 2000 Aug;6(2):395-407); mental retardation (e.g., Russo, et al., Am J Med Genet. 2000 Oct 23;94(5):376-82). Also in Mantle cell lymphoma (Xq26-q28, 5/50).(Knuutila, et al.).

SGK411, CaMKII delta2 (SEQ ID NO:12, ENCODING SEQ ID NO:69 maps to chromosomal position 4q25. This chromosomal position has been associated with the following human diseases: developmental glaucoma (Rieger syndrome, iris hypoplasia, and iridogoniodysgenesis; Craig, et al. Curr Opin Ophthalmol. 1999 Apr;10(2):126-34)

5       SGK027 (SEQ ID NO:13, ENCODING SEQ ID NO:70 maps to chromosomal position 5q11-q11.1. This chromosomal position has been associated with the following human diseases: malignant fibrous histiocyoma of bone (1/26) (Knuutila, et al.). Also with B-cell non-Hodgkin's lymphoma. (Wlodarska, et al. Cytogenet Cell Genet. 1994;65(3):179-83).

10       SGK046b (SEQ ID NO:14, ENCODING SEQ ID NO:71 maps to chromosomal position 3p24.1. This chromosomal position has been associated with the following human diseases: malignant fibrous histiocyoma of soft tissue (3p24-p26, 2/30) (Knuutila, et al.).

SGK046c (SEQ ID NO:15, ENCODING SEQ ID NO:72 maps to chromosomal position 3p24.1. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocyoma of soft tissue (3p24-p26, 2/30) (Knuutila, et al.).

15       SGK089 (SEQ ID NO:16, ENCODING SEQ ID NO:73 maps to chromosomal position 3p25.3. This chromosomal position has not been associated with human diseases.

SGK133 (SEQ ID NO:17, ENCODING SEQ ID NO:74 maps to chromosomal position 7p11.2-p21. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocyoma of bone (1/26), (Knuutila, et al.).

20       SGK004, MSK (SEQ ID NO:18, ENCODING SEQ ID NO:75 maps to chromosomal position 21q22.3. A gene which causes severe ocular alterations and occipital encephalocele (Knobloch syndrome) is mapped to 21q223 (Sertie, et al, Hum Mol Genet 1996 Jun;5(6):843-7)

SGK006 (SEQ ID NO:19, ENCODING SEQ ID NO:76 maps to chromosomal position 16q16.1. This chromosomal position has not been associated with human diseases.

25       SGK180, SNRK (SEQ ID NO:20, ENCODING SEQ ID NO:77 maps to chromosomal position 3p21.31. This chromosomal position has been associated with the following human diseases: cancer of the bladder (1/14), (Knuutila, et al.).

30       SGK386, MLCKs (SEQ ID NO:21, ENCODING SEQ ID NO:78 maps to chromosomal position 20q11.1. This chromosomal position has been associated with the following human diseases: papillary renal cell carcinoma. (Lab Invest. 1999 Mar;79(3):311-6).

SGK003 (SEQ ID NO:22, ENCODING SEQ ID NO:79 maps to chromosomal position 13q14.11. This chromosomal position has been associated with the following human diseases: fallopian tube cancer (13q14-qter; 1/12) (Knuutila, et al.).

5 SGK066 (SEQ ID NO:23, ENCODING SEQ ID NO:80 maps to chromosomal position 15q15. This chromosomal position has been associated with the following human diseases: cancer of the testis (2/11), (Knuutila, et al.).

SGK041, NKIAMRE (SEQ ID NO:24, ENCODING SEQ ID NO:81 maps to chromosomal position 5q31.1. This chromosomal position has been associated with the following human diseases: cancer of the digestive tract (5q31-qter), (Knuutila, et al.).

10 SGK112 (SEQ ID NO:25, ENCODING SEQ ID NO:82 maps to chromosomal position CHR2.

SGK038, ERK7 (SEQ ID NO:26, ENCODING SEQ ID NO:83 maps to chromosomal position na.

15 SGK158 (SEQ ID NO:27, ENCODING SEQ ID NO:84 maps to chromosomal position 1q42.11-42.2. This chromosomal position has not been associated with human diseases.

SGK429 (SEQ ID NO:28, ENCODING SEQ ID NO:85 maps to chromosomal position 7q34-35. This chromosomal position has been associated with the following human diseases: deafness ( Mustapha, et al. Eur J Hum Genet. 1998 May-Jun;6(3):245-50).

20 SGK152, SUDD (SEQ ID NO:29, ENCODING SEQ ID NO:86 maps to chromosomal position 18p11.1. This chromosomal position has not been associated with human diseases.

SGK077 (SEQ ID NO:30, ENCODING SEQ ID NO:87 maps to chromosomal position 17p13.3. This chromosomal position has been associated with the following human diseases: Lost in cervical cancer (loss of heterogeneity, Lazo, The molecular genetics of cervical carcinoma. Br J Cancer. 1999 Aug;80(12):2008-18. Review).

25 SGK093, Wnk3 (SEQ ID NO:31, ENCODING SEQ ID NO:88 maps to chromosomal position 17q21.1-2. This chromosomal position has been associated with the following human diseases: cancer of the ovary (17q21-qter, 3/47), (Knuutila, et al.).

30 SGK074 (SEQ ID NO:32, ENCODING SEQ ID NO:89 maps to chromosomal position 19p12-19q13. This chromosomal position has been associated with the following human diseases: Small cell lung cancer (2/22), (Knuutila, et al.).

SGK087 (SEQ ID NO:33, ENCODING SEQ ID NO:90 maps to chromosomal position 12p13.3. This chromosomal position has been associated with the following human diseases: hypertension (Disse-Nicodeme, et al. Am J Hum Genet. 2000 Aug;67(2):302-10).

5       SGK295, KIS (SEQ ID NO:34, ENCODING SEQ ID NO:91 maps to chromosomal position 1q23.3. This chromosomal position has been associated with the following human diseases: Hematologic neoplasms (11q23-qter; 1/1), (Knuutila, et al.).

SGK419 (SEQ ID NO:35, ENCODING SEQ ID NO:92 maps to chromosomal position 4q24. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocytoma of soft tissue (14q24-q31; 1/58), (Knuutila, et al.).

10       SGK125, MYO3A (SEQ ID NO:36, ENCODING SEQ ID NO:93 maps to chromosomal position 10p12.32. This chromosomal position has been associated with the following human diseases: Mantle cell lymphoma (10p12-p13; 2/45), (Knuutila, et al.).

SGK445 (SEQ ID NO:37, ENCODING SEQ ID NO:94 maps to chromosomal position na. (Knuutila, et al.).

15       SGK127 (SEQ ID NO:38, ENCODING SEQ ID NO:95 maps to chromosomal position 12q24.21. This chromosomal position has been associated with the following human diseases: cancer of the respiratory tract and of the female genital organs (12q24.2). (Knuutila, et al.).

SGK009, ANKRD3 (SEQ ID NO:39, ENCODING SEQ ID NO:96 maps to chromosomal position 21q22.3. A gene which causes severe ocular alterations and occipital encephalocele (Knobloch syndrome) is mapped to 21q223 (Sertie, et al, Hum Mol Genet 1996 Jun;5(6):843-7) (Knuutila, et al.).

SGK421, STK22A, TSK1 (SEQ ID NO:40, ENCODING SEQ ID NO:97 maps to chromosomal position 5q31.1. This chromosomal position has been associated with the following human diseases: Chondrosarcoma (5q31-q32; 2/45), (Knuutila, et al.).

25       SGK047 (SEQ ID NO:41, ENCODING SEQ ID NO:98 maps to chromosomal position 10p11.21. This chromosomal position has been associated with the following human diseases: Squamous cell carcinomas of the head and neck (10p11-p13; 1/30), (Knuutila, et al.).

SGK196 (SEQ ID NO:42, ENCODING SEQ ID NO:99 maps to chromosomal position na.

SGK396 (SEQ ID NO:43, ENCODING SEQ ID NO:100 maps to chromosomal position na.

SGK279, PKN (SEQ ID NO:44, ENCODING SEQ ID NO:101 maps to chromosomal position 16q22.3. This chromosomal position has been associated with the following human diseases: Diffuse large cell lymphoma of stomach (16q22-ter; 1/7), (Knuutila, et al.).

SGK037 (SEQ ID NO:45, ENCODING SEQ ID NO:102 maps to chromosomal position 13q14.12. This chromosomal position has been associated with the following human diseases: fallopian tube cancer (13q14-qter; 1/12) (Knuutila, et al.).

SGK060 (SEQ ID NO:46, ENCODING SEQ ID NO:103 maps to chromosomal position 3q22.1. This chromosomal position has been associated with the following human diseases: Mantle cell lymphoma (13q22-q32; 3/72), (Knuutila, et al.).

SGK080 (SEQ ID NO:47, ENCODING SEQ ID NO:104 maps to chromosomal position 22q11.2. This chromosomal position has been associated with the following human diseases: Non-small cell lung cancer (1/50), (Knuutila, et al.).

SGK002 (SEQ ID NO:48, ENCODING SEQ ID NO:105 maps to chromosomal position 7q32.2. This chromosomal position has been associated with the following human diseases: Non-small cell lung cancer 7q32-q35; 1/50), (Knuutila, et al.).

SGK058 (SEQ ID NO:49, ENCODING SEQ ID NO:106 maps to chromosomal position 2q21.2. This chromosomal position has been associated with the following human diseases: bladder carcinoma (12q21-q24; 1/16), (Knuutila, et al.).

SGK103 (SEQ ID NO:50, ENCODING SEQ ID NO:107 maps to chromosomal position 5p14.3. This chromosomal position has been associated with the following human diseases: Malignant fibrous histiocytoma of soft tissue (6/88), (Knuutila, et al.).

SGK035 (SEQ ID NO:51, ENCODING SEQ ID NO:108 maps to chromosomal position CHR15.

SGK075 (SEQ ID NO:52, ENCODING SEQ ID NO:109 maps to chromosomal position 2q31.1. This chromosomal position has been associated with the following human diseases: Squamous cell carcinomas of the head and neck (2q31-q33; 3/30), (Knuutila, et al.).

SGK188, EphA9 (SEQ ID NO:53, ENCODING SEQ ID NO:110 maps to chromosomal position 1p34.1-34.3. This chromosomal position has been associated with the following human diseases: cancer of the testis (1p34-pter; 1/11), (Knuutila, et al.).

5 SGK040 (SEQ ID NO:54, ENCODING SEQ ID NO:111 maps to chromosomal position 12q12. This chromosomal position has been associated with the following human diseases: Diffuse large cell lymphoma (2/66), (Knuutila, et al.).

SGK390 (SEQ ID NO:55, ENCODING SEQ ID NO:112 maps to chromosomal position 13q14.2. This chromosomal position has been associated with the following human diseases: fallopian tube cancer (13q14-qter; 1/12) (Knuutila, et al.).

10 SGK007 (SEQ ID NO:56, ENCODING SEQ ID NO:113 maps to chromosomal position 10q26.11. This chromosomal position has not been associated with human diseases. (Knuutila, et al.).

15 SGK050 (SEQ ID NO:57, ENCODING SEQ ID NO:114 maps to chromosomal position 9p13.1-p13.2. This chromosomal position has been associated with the following human diseases: Non-small cell lung cancer (1/50) and testicular cancer (4/11) (Knuutila, et al.).

### **EXAMPLE 8b: Candidate Single Nucleotide Polymorphisms (SNPs)**

#### **Materials and Methods**

20 The most common variations in human DNA are single nucleotide polymorphisms (SNPs), which occur approximately once every 100 to 300 bases. Because SNPs are expected to facilitate large-scale association genetics studies, there has recently been great interest in SNP discovery and detection. Candidate SNPs for the genes in this patent were identified by blastn searching the nucleic acid sequences against the public database of sequences containing

25 documented SNPs (dbSNP: sequence files were downloaded from <ftp://ncbi.nlm.nih.gov/SNP/human/rs-fasta/> and <ftp://ncbi.nlm.nih.gov/SNP/human/ss-fasta/> and used to create a blast database). dbSNP accession numbers for the SNP-containing sequences are given. SNPs were also identified by comparing several databases of expressed genes (dbEST, NRNA) and genomic sequence (i.e., NRNA) for single basepair mismatches. The

30 results are shown in Table 2, in the column labeled "SNPs". These are candidate SNPs – their

actual frequency in the human population was not determined. The code below is standard for representing DNA sequence:

	G	= Guanosine
5	A	= Adenosine
	T	= Thymidine
	C	= Cytidine
	R	= G or A, puRine
	Y	= C or T, pYrimidine
10	K	= G or T, Keto
	W	= A or T, Weak (2 H-bonds)
	S	= C or G, Strong (3 H-bonds)
	M	= A or C, aMino
	B	= C, G or T (i.e., not A)
15	D	= A, G or T (i.e., not C)
	H	= A, C or T (i.e., not G)
	V	= A, C or G (i.e., not T)
	N	= A, C, G or T, aNy
	X	= A, C, G or T
20		
	complementary	G A T C R Y W S K M B V D H N X
	DNA	+ - + - + - + - + - + - + - + - + - + - + - +
	strands	C T A G Y R S W M K V B H D N X

- 25 For example, if two versions of a gene exist, one with a "C" at a given position, and a second one with a "T" at the same position, then that position is represented as a Y, which means C or T. In table 2, for SGK002, the SNP column says "1165=R", which means that at position 1165, a polymorphism exists, with that position sometimes containing a G and sometimes an A (R represents A or G). SNPs may be important in identifying heritable traits associated with a gene.

30



## Results

SGK187, CRIK (SEQ ID NO:1, ENCODING SEQ ID NO:58 contains candidate single nucleotide polymorphisms at the following positions: 2874=R (ss1337340); 2883=Y (rs904655); 3327=R (ss1581624). The sequences preceding the SNPs are: 3327=R (gggttcgagagctgcagagR)  
 5 dbSNP ss1581624; 2883=Y (agagaccacagcagaagcY) dbSNP rs904655; 2874=R (gctggaagagaccacR) dbSNP ss1337340).

SGK064, GRK7 (SEQ ID NO:2, ENCODING SEQ ID NO:59 contains candidate single nucleotide polymorphisms at the following positions: 965=K; 1318=R. The sequences preceding the SNPs are: 965=K (catgaagcctgagaatgK); 1318=R (agaaaccagagcaacgcttaR).

10 SGK409, KIAA0303 (SEQ ID NO:3, ENCODING SEQ ID NO:60 contains candidate single nucleotide polymorphisms at the following positions: 6282=M ; 6327=M. The sequences preceding the SNPs are: 6327=M (ccctccaaagactaagcaccgccgaM); 6282=M (cacagtgaagcagcagccacaagccM).

SGK021 (SEQ ID NO:4, ENCODING SEQ ID NO:61 contains candidate single  
 15 nucleotide polymorphisms at the following positions: 9=S; 97=R. The sequences preceding the SNPs are: 9=S (atgggagcS); 97=R (agccattgggaaaggcR).

SGK410 (SEQ ID NO:5, ENCODING SEQ ID NO:62 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK069 (SEQ ID NO:6, ENCODING SEQ ID NO:63 contains candidate single  
 20 nucleotide polymorphisms at the following positions: 1180=S (ss1317629); 210=Y (ss1688813). The sequences preceding the SNPs are: 1180=S (ctggttcggcctggccS) dbSNP ss1317629; 210=Y (tgcaggacggggaY) dbSNP ss1688813.

SGK110 (SEQ ID NO:7, ENCODING SEQ ID NO:64 contains candidate single nucleotide polymorphisms at the following positions: 597=R (rs654439); 252=Y (ss661406).  
 25 The sequences preceding the SNPs are: 597=R (aacgtgctggtcttcgaccR) dbSNP rs654439; 252=Y (ccgcgtgctcctY) dbSNP ss661406.

SGK053, CKLiK (SEQ ID NO:8, ENCODING SEQ ID NO:65 contains candidate single nucleotide polymorphisms at the following positions: 605=Y; 509=M. The sequences preceding the SNPs are: 605=Y (aaccttacagcaaagccgY); 509=M (atcagtgactttggattgtcaaM).

SGK124 (SEQ ID NO:9, ENCODING SEQ ID NO:66 contains candidate single nucleotide polymorphisms at the following positions: 188=S; 333=Y. The sequences preceding the SNPs are: 188=S (tccagatcgtgcaactgctgtggS); 333=Y (tggccgtgctggagccctaY).

5     SGK254, CAMKKa (SEQ ID NO:10, ENCODING SEQ ID NO:67 contains candidate single nucleotide polymorphisms at the following positions: 555=R (ss84265); 1148=R. The sequences preceding the SNPs are: 555=R (gtcgccctccccR) dbSNP ss84265; 1148=R (tggtgttctctgaggR).

10    SGK297, CaMKIb2 (SEQ ID NO:11, ENCODING SEQ ID NO:68 contains candidate single nucleotide polymorphisms at the following positions: 77=K. The sequences preceding the SNPs are: 77=K (ggctcggctcgggtgcctK).

SGK411, CaMKII delta2 (SEQ ID NO:12, ENCODING SEQ ID NO:69 contains candidate single nucleotide polymorphisms at the following positions: 15=M; 1387=S (ss1531091). The sequences preceding the SNPs are: 15=M; atggcttcgaccacM; 1387=S (ccgggatggaaagtggS) dbSNP ss1531091.

15    SGK027 (SEQ ID NO:13, ENCODING SEQ ID NO:70 contains candidate single nucleotide polymorphisms at the following positions: 45=Y. The sequences preceding the SNPs are: 45=Y (tgaatggaggtggcctggtgaaccccccaY).

SGK046b (SEQ ID NO:14, ENCODING SEQ ID NO:71 contains candidate single nucleotide polymorphisms at the following positions: none detected.

20    SGK046c (SEQ ID NO:15, ENCODING SEQ ID NO:72 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK089 (SEQ ID NO:16, ENCODING SEQ ID NO:73 contains candidate single nucleotide polymorphisms at the following positions: none detected.

25    SGK133 (SEQ ID NO:17, ENCODING SEQ ID NO:74 contains candidate single nucleotide polymorphisms at the following positions: 2003=S;1673=S. The sequences preceding the SNPs are: 2003=S (gcttcagggccgagtagaaggccagtS); 1673=S (cacccccagccccggcgS).

SGK004, MSK (SEQ ID NO:18, ENCODING SEQ ID NO:75 contains candidate single nucleotide polymorphisms at the following positions: 1853=Y (ss571239). The sequences preceding the SNPs are: 1853=Y (aggtgtgccaggccccY) dbSNP ss571239.

SGK006 (SEQ ID NO:19, ENCODING SEQ ID NO:76 contains candidate single nucleotide polymorphisms at the following positions: 4=S (ss1609852). The sequences preceding the SNPs are: 4=S (Sctctgtatttcaggaagctct) dbSNP ss1609852.

5 SGK180, SNRK (SEQ ID NO:20, ENCODING SEQ ID NO:77 contains candidate single nucleotide polymorphisms at the following positions: 1817=S. The sequences preceding the SNPs are: 1817=S (agccccagtgagaacaatgctggtgggS).

SGK386, MLCKs (SEQ ID NO:21, ENCODING SEQ ID NO:78 contains candidate single nucleotide polymorphisms at the following positions: 835=M. The sequences preceding the SNPs are: 835=M (ttttgatgattgcccgccaM).

10 SGK003 (SEQ ID NO:22, ENCODING SEQ ID NO:79 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK066 (SEQ ID NO:23, ENCODING SEQ ID NO:80 contains candidate single nucleotide polymorphisms at the following positions: none detected.

15 SGK041, NKIAMRE (SEQ ID NO:24, ENCODING SEQ ID NO:81 contains candidate single nucleotide polymorphisms at the following positions: 1033=R;1284=R; 1181=Y. The sequences preceding the SNPs are: 1033=R (ctaagtagttcagtttgggaR);1284=R (tccacattgcggaggttctgtR); 1181=Y (gcaaatgaaatgttcacctaY).

SGK112 (SEQ ID NO:25, ENCODING SEQ ID NO:82 contains candidate single nucleotide polymorphisms at the following positions: none detected.

20 SGK038, ERK7 (SEQ ID NO:26, ENCODING SEQ ID NO:83 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK158 (SEQ ID NO:27, ENCODING SEQ ID NO:84 contains candidate single nucleotide polymorphisms at the following positions: 1752=Y (ss1529336). The sequences preceding the SNPs are: 1752=Y (atcctggttctgtgcctgaggagctgttY) dbSNP ss1529336.

25 SGK429 (SEQ ID NO:28, ENCODING SEQ ID NO:85 contains candidate single nucleotide polymorphisms at the following positions: 196=R; 919=Y (ss1549835); 1865=Y (ss1517749). The sequences preceding the SNPs are: 196=R (cccctgacgttctgR); 919=Y (aggccaccaacctcatcY) dbSNP ss1549835; 1865=Y (CCTCCTCACGGGCCY) dbSNP ss1517749.

SGK152, SUDD (SEQ ID NO:29, ENCODING SEQ ID NO:86 contains candidate single nucleotide polymorphisms at the following positions: 972=M. The sequences preceding the SNPs are: 972=M (AAACTAAATCCACGTAAGATCM).

5 SGK077 (SEQ ID NO:30, ENCODING SEQ ID NO:87 contains candidate single nucleotide polymorphisms at the following positions: 390=Y (ss1658885); 611=R (ss1629760); 985=Y (ss1629759). The sequences preceding the SNPs are: 390=Y (agcacaccctgY) dbSNP ss1658885; 611=R (cagccgtcccgagcgR) dbSNP ss1629760; 985=Y (ttgtgccaaggggaaY) dbSNP ss1629759.

10 SGK093, Wnk3 (SEQ ID NO:31, ENCODING SEQ ID NO:88 contains candidate single nucleotide polymorphisms at the following positions: 3279 = K; 4078=M. The sequences preceding the SNPs are: 3279=K (gcccccaaccctgagccaK); 4078=M (caaaataatcagcaacattM) .

SGK074 (SEQ ID NO:32, ENCODING SEQ ID NO:89 contains candidate single nucleotide polymorphisms at the following positions: none detected.

15 SGK087 (SEQ ID NO:33, ENCODING SEQ ID NO:90 contains candidate single nucleotide polymorphisms at the following positions: 269=R (ss88136). The sequences preceding the SNPs are: 269=R (acttcacgagcagtcaggcR) dbSNP ss88136.

SGK295, KIS (SEQ ID NO:34, ENCODING SEQ ID NO:91 contains candidate single nucleotide polymorphisms at the following positions: 355=W. The sequences preceding the SNPs are: 355=W (ctggatgtcagtggtW).

20 SGK419 (SEQ ID NO:35, ENCODING SEQ ID NO:92 contains candidate single nucleotide polymorphisms at the following positions: none

SGK125, MYO3A (SEQ ID NO:36, ENCODING SEQ ID NO:93 contains candidate single nucleotide polymorphisms at the following positions: 3145=Y; 3204=Y. The sequences preceding the SNPs are: 3145=Y (attatcacgtggagcagttaaatY); 3204=Y (gattcaagcttgtgtcagagcattY).

SGK445 (SEQ ID NO:37, ENCODING SEQ ID NO:94 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK127 (SEQ ID NO:38, ENCODING SEQ ID NO:95 contains candidate single nucleotide polymorphisms at the following positions: 501=S (ss2005786). The sequences preceding the SNPs are: 501=S (tccagtggccacS) dbSNP ss2005786.

SGK009, ANKRD3 (SEQ ID NO:39, ENCODING SEQ ID NO:96 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK421, STK22A, TSK1 (SEQ ID NO:40, ENCODING SEQ ID NO:97 contains candidate single nucleotide polymorphisms at the following positions: 510=M (ss2055126);  
 5 279=R (ss2055125). The sequences preceding the SNPs are: 510=M (ACAGTGGTCGAATGGCM) dbSNP ss2055126; 279=R (gtcatggagctcgcR) dbSNP ss2055125.

SGK047 (SEQ ID NO:41, ENCODING SEQ ID NO:98 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK196 (SEQ ID NO:42, ENCODING SEQ ID NO:99 contains candidate single  
 10 nucleotide polymorphisms at the following positions: 99=R. The sequences preceding the SNPs are: 99=R (ctgctgatcatggccctgatgaatactctR).

SGK396 (SEQ ID NO:43, ENCODING SEQ ID NO:100 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK279, PKN (SEQ ID NO:44, ENCODING SEQ ID NO:101 contains candidate single  
 15 nucleotide polymorphisms at the following positions: 665=Y. The sequences preceding the SNPs are: 665=Y (CCTGAGGTGTGCCAGGY).

SGK037 (SEQ ID NO:45, ENCODING SEQ ID NO:102 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK060 (SEQ ID NO:46, ENCODING SEQ ID NO:103 contains candidate single  
 20 nucleotide polymorphisms at the following positions: 1463=W. The sequences preceding the SNPs are: 1463=W (tgatgcatttgattcctattgtgW).

SGK080 (SEQ ID NO:47, ENCODING SEQ ID NO:104 contains candidate single nucleotide polymorphisms at the following positions: 1159=R (ss1367671); 422=R (ss1855009).  
 The sequences preceding the SNPs are: 1159=R (aagttcatttcagtggR) dbSNP ss1367671; 422=R  
 25 (tactgtagtgcgtcgggR) dbSNP ss1855009.

SGK002 (SEQ ID NO:48, ENCODING SEQ ID NO:105 contains candidate single nucleotide polymorphisms at the following positions: 1165=R; 983=M. The sequences preceding the SNPs are: 1165=R (tgaggccatctttggccR); 983=M, 1753=R. (GGGGTGAGCGGCCAGCTCATM); 443=K, (ACCTGGTGGACCTGCAGAAK); 1245=Y

(GCGCCCCAACAGCGGTY); 1261=K (GGGATGGACAGCCK), 1753 =R (tggtccctgccggcgtgcctctR).

SGK058 (SEQ ID NO:49, ENCODING SEQ ID NO:106 contains candidate single nucleotide polymorphisms at the following positions: none detected.

5       SGK103 (SEQ ID NO:50, ENCODING SEQ ID NO:107 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK035 (SEQ ID NO:51, ENCODING SEQ ID NO:108 contains candidate single nucleotide polymorphisms at the following positions: 2273=Y. The sequences preceding the SNPs are: 273=Y (cagttttcactgctacY).

10       SGK075 (SEQ ID NO:52, ENCODING SEQ ID NO:109 contains candidate single nucleotide polymorphisms at the following positions: 889=R. The sequences preceding the SNPs are: 889=R (tttctgcaaaaacagctggccR).

SGK188, EphA9 (SEQ ID NO:53, ENCODING SEQ ID NO:110 contains candidate single nucleotide polymorphisms at the following positions: 2104=Y (ss1986120); 2319=R. The sequences preceding the SNPs are: 2319=R (gtggctgggcaactgatggR); 2104=Y (gtagccgtgcaY) dbSNP ss1986120.

SGK040 (SEQ ID NO:54, ENCODING SEQ ID NO:111 contains candidate single nucleotide polymorphisms at the following positions: 1869 = R, 1004 = Y. The sequences preceding the SNPs are: 1869 = R (ggttggtgccccatggcctatggttgaR); 1004 = Y (gacttgatttggctgaccY).

SGK390 (SEQ ID NO:55, ENCODING SEQ ID NO:112 contains candidate single nucleotide polymorphisms at the following positions: 1314=R. The sequences preceding the SNPs are: 1314=R (gggaggttcataatgacgatgacacccaR).

SGK007 (SEQ ID NO:56, ENCODING SEQ ID NO:113 contains candidate single nucleotide polymorphisms at the following positions: none detected.

SGK050 (SEQ ID NO:57, ENCODING SEQ ID NO:114 contains candidate single nucleotide polymorphisms at the following positions: none detected.

### 30    EXAMPLE 9: Demonstration Of Gene Amplification By Southern Blotting

#### Materials and Methods

Nylon membranes are purchased from Boehringer Mannheim. Denaturing solution contains 0.4 M NaOH and 0.6 M NaCl. Neutralization solution contains 0.5 M Tris-HCL, pH 7.5 and 1.5 M NaCl. Hybridization solution contains 50% formamide, 6X SSPE, 2.5X Denhardt's solution, 0.2 mg/mL denatured salmon DNA, 0.1 mg/mL yeast tRNA, and 0.2 % sodium dodecyl sulfate. Restriction enzymes are purchased from Boehringer Mannheim. Radiolabeled probes are prepared using the Prime-it II kit by Stratagene. The beta actin DNA fragment used for a probe template is purchased from Clontech.

Genomic DNA is isolated from a variety of tumor cell lines (such as MCF-7, MDA-MB-231, Calu-6, A549, HCT-15, HT-29, Colo 205, LS-180, DLD-1, HCT-116, PC3, CAPAN-2, MIA-PaCa-2, PANC-1, AsPc-1, BxPC-3, OVCAR-3, SKOV3, SW 626 and PA-1, and from two normal cell lines.

A 10 µg aliquot of each genomic DNA sample is digested with EcoR I restriction enzyme and a separate 10 µg sample is digested with Hind III restriction enzyme. The restriction-digested DNA samples are loaded onto a 0.7% agarose gel and, following electrophoretic separation, the DNA is capillary-transferred to a nylon membrane by standard methods (Sambrook, J. *et al* (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory).

#### EXAMPLE 10: Detection Of Protein-Protein Interaction Through Phage Display

##### Materials And Methods

Phage display provides a method for isolating molecular interactions based on affinity for a desired bait. cDNA fragments cloned as fusions to phage coat proteins are displayed on the surface of the phage. Phage(s) interacting with a bait are enriched by affinity purification and the insert DNA from individual clones is analyzed.

##### T7 Phage Display Libraries

All libraries were constructed in the T7Select1-1b vector (Novagen) according to the manufacturer's directions.

##### Bait Presentation

Protein domains to be used as baits are generated as C-terminal fusions to GST and expressed in *E. coli*. Peptides are chemically synthesized and biotinylated at the N-terminus using a long chain spacer biotin reagent.

5        Selection

Aliquots of refreshed libraries ( $10^{10}$ - $10^{12}$  pfu) supplemented with PanMix and a cocktail of *E. coli* inhibitors (Sigma P-8465) are incubated for 1-2 hrs at room temperature with the immobilized baits. Unbound phage is extensively washed (at least 4 times) with wash buffer.

After 3-4 rounds of selection, bound phage is eluted in 100  $\mu$ L of 1% SDS and plated on  
10 agarose plates to obtain single plaques.

Identification of insert DNAs

Individual plaques are picked into 25  $\mu$ L of 10 mM EDTA and the phage is disrupted by heating at 70 °C for 10 min. 2  $\mu$ L of the disrupted phage are added to 50  $\mu$ L PCR reaction mix.

15 The insert DNA is amplified by 35 rounds of thermal cycling (94 °C, 50 sec; 50 °C, 1min; 72 °C, 1min).

Composition of Buffer

10x PanMix  
20 5% Triton X-100  
10% non-fat dry milk (Carnation)  
10 mM EGTA  
250 mM NaF  
250  $\mu$ g/mL Heparin (sigma)  
25 250  $\mu$ g/mL sheared, boiled salmon sperm DNA (sigma)  
0.05% Na azide  
Prepared in PBS

Wash Buffer

30 PBS supplemented with:  
0.5% NP-40



25 µl g/mL heparin

PCR reaction mix

- 1.0 mL 10x PCR buffer (Perkin-Elmer, with 15 mM Mg)
- 0.2 mL each dNTPs (10 mM stock)
- 5 0.1 mL T7UP primer (15 pmol/µL) GGAGCTGTCGTATTCCAGTC
- 0.1 mL T7DN primer (15 pmol/µL) AACCCCTCAAGACCCGTTTAG
- 0.2 mL 25 mM MgCl<sub>2</sub> or MgSO<sub>4</sub> to compensate for EDTA
- Q.S. to 10 mL with distilled water
- Add 1 unit of Taq polymerase per 50 µL reaction
- 10 LIBRARY: T7 Select1-H441

#### EXAMPLE 11: FLK-1

15 An ELISA assay was conducted to measure the kinase activity of the FLK-1 receptor and more specifically, the inhibition or activation of TK activity on the FLK-1 receptor. Specifically, the following assay was conducted to measure kinase activity of the FLK-1 receptor in cells genetically engineered to express Flk-1.

#### Materials and Reagents

The following reagents and supplies were used:

- 20 1. Corning 96-well ELISA plates (Corning Catalog No. 25805-96);
2. Cappel goat anti-rabbit IgG (catalog no. 55641);
3. PBS (Gibco Catalog No. 450-1300EB);
4. TBSW Buffer (50 mM Tris (pH 7.2), 150 mM NaCl and 0.1% Tween-20);
5. Ethanolamine stock (10% ethanolamine (pH 7.0), stored at 4 °C);
- 25 6. HNTG buffer (20 mM HEPES buffer (pH 7.5), 150 mM NaCl, 0.2% Triton X-100, and 10% glycerol);
7. EDTA (0.5 M (pH 7.0) as a 100X stock);
8. Sodium orthovanadate (0.5 M as a 100X stock);
9. Sodium pyrophosphate (0.2 M as a 100X stock);
- 30 10. NUNC 96 well V bottom polypropylene plates (Applied Scientific Catalog No. AS-72092);

11. NIH3T3 C7#3 Cells (FLK-1 expressing cells);
12. DMEM with 1X high glucose L-Glutamine (catalog No. 11965-050);
13. FBS, Gibco (catalog no. 16000-028);
14. L-glutamine, Gibco (catalog no. 25030-016);
- 5 15. VEGF, PeproTech, Inc. (catalog no. 100-20) (kept as 1  $\mu$ g/100  $\mu$ l stock in Milli-Q dH<sub>2</sub>O and stored at -20 °C);
16. Affinity purified anti-FLK-1 antiserum;
17. UB40 monoclonal antibody specific for phosphotyrosine (see, Fendley, *et al.*, 1990, *Cancer Research* 50:1550-1558);
- 10 18. EIA grade Goat anti-mouse IgG-POD (BioRad catalog no. 172-1011);
19. 2,2-azino-bis(3-ethylbenz-thiazoline-6-sulfonic acid (ABTS) solution (100 mM citric acid (anhydrous), 250 mM Na<sub>2</sub>HPO<sub>4</sub> (pH 4.0), 0.5 mg/ml ABTS (Sigma catalog no. A-1888)), solution should be stored in dark at 4 °C until ready for use;
20. H<sub>2</sub>O<sub>2</sub> (30% solution) (Fisher catalog no. H325);
- 15 21. ABTS/H<sub>2</sub>O<sub>2</sub> (15 ml ABTS solution, 2  $\mu$ l H<sub>2</sub>O<sub>2</sub>) prepared 5 minutes before use and left at room temperature;
22. 0.2 M HCl stock in H<sub>2</sub>O;
23. dimethylsulfoxide (100%) (Sigma Catalog No. D-8418); and
24. Trypsin-EDTA (Gibco BRL Catalog No. 25200-049).

20

#### Protocol

The following protocol was used for conducting the assay:

1. Coat Corning 96-well ELISA plates with 1.0  $\mu$ g per well Cappel Anti-rabbit IgG antibody in 0.1 M Na<sub>2</sub>CO<sub>3</sub> pH 9.6. Bring final volume to 150  $\mu$ l per well. Coat plates overnight  
25 at 4 °C. Plates can be kept up to two weeks when stored at 4 °C.
2. Grow cells in Growth media (DMEM, supplemented with 2.0 mM L-Glutamine, 10% FBS) in suitable culture dishes until confluent at 37 °C, 5% CO<sub>2</sub>.
3. Harvest cells by trypsinization and seed in Corning 25850 polystyrene 96-well round bottom cell plates, 25,000 cells/well in 200  $\mu$ l of growth media.
- 30 4. Grow cells at least one day at 37 °C, 5% CO<sub>2</sub>.

5. Wash cells with D-PBS 1X.
6. Add 200  $\mu$ l/well of starvation media (DMEM, 2.0 mM l-Glutamine, 0.1% FBS).  
Incubate overnight at 37 °C, 5% CO<sub>2</sub>.
7. Dilute Compounds 1:20 in polypropylene 96 well plates using starvation media.
- 5 Dilute dimethylsulfoxide 1:20 for use in control wells.
8. Remove starvation media from 96 well cell culture plates and add 162  $\mu$ l of fresh starvation media to each well.
9. Add 18  $\mu$ l of 1:20 diluted Compound dilution (from step 7) to each well plus the 1:20 dimethylsulfoxide dilution to the control wells ( $\pm$  VEGF), for a final dilution of 1:200 after  
10 cell stimulation. Final dimethylsulfoxide is 0.5%. Incubate the plate at 37 °C, 5% CO<sub>2</sub> for two hours.
10. Remove unbound antibody from ELISA plates by inverting plate to remove liquid. Wash 3 times with TBSW + 0.5% ethanolamine, pH 7.0. Pat the plate on a paper towel to remove excess liquid and bubbles.
- 15 11. Block plates with TBSW + 0.5% Ethanolamine, pH 7.0, 150  $\mu$ l per well. Incubate plate thirty minutes while shaking on a microtiter plate shaker.
12. Wash plate 3 times as described in step 10.
13. Add 0.5  $\mu$ g/well affinity purified anti-FLU-1 polyclonal rabbit antiserum. Bring  
final volume to 150  $\mu$ l/well with TBSW + 0.5% ethanolamine pH 7.0. Incubate plate for thirty  
20 minutes while shaking.
14. Add 180  $\mu$ l starvation medium to the cells and stimulate cells with 20  $\mu$ l/well 10.0 mM sodium ortho vanadate and 500 ng/ml VEGF (resulting in a final concentration of 1.0 mM sodium ortho vanadate and 50 ng/ml VEGF per well) for eight minutes at 37 °C, 5% CO<sub>2</sub>.  
Negative control wells receive only starvation medium.
- 25 15. After eight minutes, media should be removed from the cells and washed one time with 200  $\mu$ l/well PBS.
16. Lyse cells in 150  $\mu$ l/well HNTG while shaking at room temperature for five minutes. HNTG formulation includes sodium ortho vanadate, sodium pyrophosphate and EDTA.

17. Wash ELISA plate three times as described in step 10.
18. Transfer cell lysates from the cell plate to ELISA plate and incubate while shaking for two hours. To transfer cell lysate pipette up and down while scrapping the wells.
19. Wash plate three times as described in step 10.
- 5 20. Incubate ELISA plate with 0.02 µg/well UB40 in TBSW + 0.5% ethanolamine. Bring final volume to 150 µl/well. Incubate while shaking for 30 minutes.
21. Wash plate three times as described in step 10.
22. Incubate ELISA plate with 1:10,000 diluted EIA grade goat anti-mouse IgG conjugated horseradish peroxidase in TBSW + 0.5% ethanolamine, pH 7.0. Bring final volume  
10 to 150 µl/well. Incubate while shaking for thirty minutes.
23. Wash plate as described in step 10.
24. Add 100 µl of ABTS/H<sub>2</sub>O<sub>2</sub> solution to well. Incubate ten minutes while shaking.
25. Add 100 µl of 0.2 M HCl for 0.1 M HCl final to stop the color development  
15 reaction. Shake 1 minute at room temperature. Remove bubbles with slow stream of air and read the ELISA plate in an ELISA plate reader at 410 nm.

#### EXAMPLE 12: HER-2 ELISA

- Assay 1: EGF Receptor-HER2 Chimeric Receptor Assay In Whole Cells.
- 20 HER2 kinase activity in whole EGFR-NIH3T3 cells was measured as described below:

#### Materials and Reagents

The following materials and reagents were used to conduct the assay:

1. EGF: stock concentration: 16.5 ILM; EGF 201, TOYOBO, Co., Ltd. Japan.
- 25 2. 05-101 (UBI) (a monoclonal antibody recognizing an EGFR extracellular domain).
3. Anti-phosphotyrosine antibody (anti-Ptyr) (polyclonal) (*see, Fendley, et al., supra*).
4. Detection antibody: Goat anti-rabbit IgG horse radish peroxidase conjugate,  
30 TAGO, Inc., Burlingame, CA.
5. TBST buffer:

- |    |                                  |           |
|----|----------------------------------|-----------|
|    | Tris-HCl, pH 7.2                 | 50 mM     |
|    | NaCl                             | 150 mM    |
|    | Triton X-100                     | 0.1       |
| 5  | 6. HNTG 5X stock:                |           |
|    | HEPES                            | 0.1 M     |
|    | NaCl                             | 0.75 M    |
|    | Glycerol                         | 50%       |
|    | Triton X-100                     | 1.0%      |
| 10 | 7. ABTS stock:                   |           |
|    | Citric Acid                      | 100 mM    |
|    | Na <sub>2</sub> HPO <sub>4</sub> | 250 mM    |
|    | HCl, conc.                       | 0.5 pM    |
|    | ABTS*                            | 0.5 mg/ml |
- \* (2,2'-azinobis(3-ethylbenzthiazolinesulfonic acid)). Keep solution in dark at 4 °C until
- 15 use.
8. Stock reagents of:
- |  |  |
|--|--|
|  | EDTA 100 mM pH 7.0                                     |
|  | Na <sub>3</sub> VO <sub>4</sub> 0.5 M                  |
|  | Na <sub>4</sub> (P <sub>2</sub> O <sub>7</sub> ) 0.2 M |

20

Protocol

The following protocol was used:

A. Pre-coat ELISA Plate

1. Coat ELISA plates (Corning, 96 well, Cat. #25805-96) with 05-101 antibody at  
25 0.5 g per well in PBS, 100 µl final volume/well, and store overnight at 4 °C. Coated plates are  
good for up to 10 days when stored at 4 °C.
2. On day of use, remove coating buffer and replace with 100 µl blocking buffer  
(5% Carnation Instant Non-Fat Dry Milk in PBS). Incubate the plate, shaking, at room  
temperature (about 23°C to 25°C) for 30 minutes. Just prior to use, remove blocking buffer and  
30 wash plate 4 times with TBST buffer.

B. Seeding Cells

1. An NIH3T3 cell line overexpressing a chimeric receptor containing the EGFR extracellular domain and intracellular HER2 kinase domain can be used for this assay.
2. Choose dishes having 80-90% confluence for the experiment. Trypsinize cells and stop reaction by adding 10% fetal bovine serum. Suspend cells in DMEM medium (10% CS DMEM medium) and centrifuge once at 1500 rpm, at room temperature for 5 minutes.
3. Resuspend cells in seeding medium (DMEM, 0.5% bovine serum), and count the cells using trypan blue. Viability above 90% is acceptable. Seed cells in DMEM medium (0.5% bovine serum) at a density of 10,000 cells per well, 100  $\mu$ l per well, in a 96 well microtiter plate.
- 10 Incubate seeded cells in 5% CO<sub>2</sub> at 37 °C for about 40 hours.

C. Assay Procedures

1. Check seeded cells for contamination using an inverted microscope. Dilute drug stock (10 mg/ml in DMSO) 1:10 in DMEM medium, then transfer 5  $\mu$ l to a TBST well for a final drug dilution of 1:200 and a final DMSO concentration of 1%. Control wells receive DMSO alone. Incubate in 5% CO<sub>2</sub> at 37 °C for two hours.
2. Prepare EGF ligand: dilute stock EGF in DMEM so that upon transfer of 10  $\mu$ l dilute EGF (1:12 dilution), 100 nM final concentration is attained.
3. Prepare fresh HNTG\* sufficient for 100  $\mu$ l per well; and place on ice.  
HNTG\* (10 ml):
 

HNTG stock	2.0 ml
milli-Q H <sub>2</sub> O	7.3 ml
EDTA, 100 mM, pH 7.0	0.5 ml
Na <sub>3</sub> VO <sub>4</sub> , 0.5 M	0.1 ml
Na <sub>4</sub> (P <sub>2</sub> O <sub>7</sub> ), 0.2 M	0.1 ml
- 25 4. After 120 minutes incubation with drug, add prepared SGF ligand to cells, 10  $\mu$ l per well, to a final concentration of 100 nM. Control wells receive DMEM alone. Incubate, shaking, at room temperature, for 5 minutes.
5. Remove drug, EGF, and DMEM. Wash cells twice with PBS. Transfer HNTG\* to cells, 100  $\mu$ l per well. Place on ice for 5 minutes. Meanwhile, remove blocking buffer from other ELISA plate and wash with TBST as described above.
- 30

6. With a pipette tip securely fitted to a micropipettor, scrape cells from plate and homogenize cell material by repeatedly aspirating and dispensing the HNTG\* lysis buffer. Transfer lysate to a coated, blocked, and washed ELISA plate. Incubate shaking at room temperature for one hour.
- 5 7. Remove lysate and wash 4 times with TBST. Transfer freshly diluted anti-Ptyr antibody to ELISA plate at 100 µl per well. Incubate shaking at room temperature for 30 minutes in the presence of the anti-Ptyr antiserum (1:3000 dilution in TBST).
8. Remove the anti-Ptyr antibody and wash 4 times with TBST. Transfer the freshly diluted TAGO anti-rabbit IgG antibody to the ELISA plate at 100 µl per well. Incubate shaking at room temperature for 30 minutes (anti-rabbit IgG antibody: 1:3000 dilution in TBST).
- 10 9. Remove TAGO detection antibody and wash 4 times with TBST. Transfer freshly prepared ABTS/H<sub>2</sub>O<sub>2</sub> solution to ELISA plate, 100 µl per well. Incubate shaking at room temperature for 20 minutes. (ABTS/H<sub>2</sub>O<sub>2</sub> solution: 1.0 µl 30% H<sub>2</sub>O<sub>2</sub> in 10 ml ABTS stock).
10. Stop reaction by adding 50 µl 5 N H<sub>2</sub>SO<sub>4</sub> (optional), and determine O.D. at 410 nm.
- 15 11. The maximal phosphotyrosine signal is determined by subtracting the value of the negative controls from the positive controls. The percent inhibition of phosphotyrosine content for extract-containing wells is then calculated, after subtraction of the negative controls.

20 EXAMPLE 13: PDGF-R ELISA

All cell culture media, glutamine, and fetal bovine serum were purchased from Gibco Life Technologies (Grand Island, NY) unless otherwise specified. All cells were grown in a humid atmosphere of 90-95% air and 5-10% CO<sub>2</sub> at 37 °C. All cell lines were routinely subcultured twice a week and were negative for mycoplasma as determined by the Mycotect method (Gibco).

25

For ELISA assays, cells (U124 2, obtained from Joseph Schlessinger, NYU) were grown to 80-90% confluency in growth medium (MEM with 10% FBS, NEAA, 1 mM NaPyr and 2 mM GLN) and seeded in 96-well tissue culture plates in 0.5% serum at 25,000 to 30,000 cells per well. After overnight incubation in 0.5% serum-containing medium, cells were changed to serum-free medium and treated with test compound for 2 hr in a 5% CO<sub>2</sub>, 37 °C incubator. Cells

30

were then stimulated with ligand for 5-10 minute followed by lysis with HNTG (20 mM Hepes, 150 mM NaCl, 10% glycerol, 5 mM EDTA, 5 mM Na<sub>3</sub>VO<sub>4</sub>, 0.2% Triton X-100, and 2 mM NaPyr). Cell lysates (0.5 mg/well in PBS) were transferred to ELISA plates previously coated with receptor-specific antibody and which had been blocked with 5% milk in TBST (50 mM Tris-HCl pH 7.2, 150 mM NaCl and 0.1% Triton X-100) at room temperature for 30 min. Lysates were incubated with shaking for 1 hour at room temperature. The plates were washed with TBST four times and then incubated with polyclonal anti-phosphotyrosine antibody at room temperature for 30 minutes. Excess anti-phosphotyrosine antibody was removed by rinsing the plate with TBST four times. Goat anti-rabbit IgG antibody was added to the ELISA plate for 30 min at room temperature followed by rinsing with TBST four more times. ABTS (100 mM citric acid, 250 mM Na<sub>2</sub>HPO<sub>4</sub> and 0.5 mg/ml 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid)) plus H<sub>2</sub>O<sub>2</sub> (1.2 ml 30% H<sub>2</sub>O<sub>2</sub> to 10 ml ABTS) was added to the ELISA plates to start color development. Absorbance at 410 nm with a reference wavelength of 630 nm was recorded about 15 to 30 min after ABTS addition.

#### EXAMPLE 14: IGF-I Receptor ELISA

The following protocol may be used to measure phosphotyrosine level on IGF-I receptor, which indicates IGF-I receptor tyrosine kinase activity.

##### Materials and Reagents

The following materials and reagents were used:

1. The cell line used in this assay is 3T3/IGF-1R, a cell line genetically engineered to overexpresses IGF-1 receptor.
2. NIH3T3/IGF-1R is grown in an incubator with 5% CO<sub>2</sub> at 37 °C. The growth media is DMEM + 10% FBS (heat inactivated)+ 2 mM L-glutamine.
3. Affinity purified anti-IGF-1R antibody 17-69.
4. D-PBS:

KH <sub>2</sub> PO <sub>4</sub>	0.20 g/L
K <sub>2</sub> HPO <sub>4</sub>	2.16 g/L
KCl	0.20 g/L



NaCl 8.00 g/L(pH 7.2)

5. Blocking Buffer: TBST plus 5% Milk (Carnation Instant Non-Fat Dry Milk).

6. TBST buffer:

5 Tris-HCl 50 mM  
NaCl 150 mM (pH 7.2/HCl 10 N)  
Triton X-100 0.1%

Stock solution of TBS (10X) is prepared, and Triton X-100 is added to the buffer during dilution.

7. HNTG buffer:

10 HEPES 20 mM  
NaCl 150 mM (pH 7.2/HCl 1 N)  
Glycerol 10%  
Triton X-100 0.2%

Stock solution (5X) is prepared and kept at 4 °C.

15 8. EDTA/HCl: 0.5 M pH 7.0 (NaOH) as 100X stock.

9. Na<sub>3</sub>VO<sub>4</sub> : 0.5 M as 100X stock and aliquots are kept in -80 °C.

10. Na<sub>4</sub> P<sub>2</sub>O<sub>7</sub>: 0.2 M as 100X stock.

11. Insulin-like growth factor-1 from Promega (Cat# G5111).

12. Rabbit polyclonal anti-phosphotyrosine antiserum.

20 13. Goat anti-rabbit IgG, POD conjugate (detection antibody), Tago (Cat. No. 4 520, Lot No. 1802): Tago, Inc., Burlingame, CA.

14. ABTS (2,2'-azinobis(3-ethylbenzthiazolinesulfonic acid)) solution:

25 Citric acid 100 mM  
Na<sub>2</sub>HPO<sub>4</sub> 250 mM (pH 4.0/1 N HCl)  
ABTS 0.5 mg/ml

ABTS solution should be kept in dark and 4 °C. The solution should be discarded when it turns green.

15. Hydrogen Peroxide: 30% solution is kept in the dark and at 4 °C.

30 Protocol

All the following steps are conducted at room temperature unless it is specifically indicated. All ELISA plate washings are performed by rinsing the plate with tap water three times, followed by one TBST rinse. Pat plate dry with paper towels.

A. Cell Seeding:

- 5        1. The cells, grown in tissue culture dish (Corning 25020-100) to 80-90% confluence, are harvested with Trypsin-EDTA (0.25%, 0.5 ml/D-100, GIBCO).
2. Resuspend the cells in fresh DMEM + 10% FBS + 2 mM L-Glutamine, and transfer to 96-well tissue culture plate (Corning, 25806-96) at 20,000 cells/well (100  $\mu$ l/well). Incubate for 1 day then replace medium to serum-free medium (90/ $\mu$ l) and incubate in 5% CO<sub>2</sub>
- 10      and 37 °C overnight.

B. ELISA Plate Coating and Blocking:

1. Coat the ELISA plate (Corning 25805-96) with Anti-IGF-1R Antibody at 0.5  $\mu$ g/well in 100  $\mu$ l PBS at least 2 hours.
2. Remove the coating solution, and replace with 100  $\mu$ l Blocking Buffer, and shake
- 15      for 30 minutes. Remove the blocking buffer and wash the plate just before adding lysate.

C. Assay Procedures:

1. The drugs are tested in serum-free condition.
2. Dilute drug stock (in 100% DMSO) 1:10 with DMEM in 96-well poly-propylene plate, and transfer 10  $\mu$ l/well of this solution to the cells to achieve final drug dilution 1:100, and
- 20      final DMSO concentration of 1.0%. Incubate the cells in 5% CO<sub>2</sub> at 37 °C for 2 hours.

3. Prepare fresh cell lysis buffer (HNTG\*)

	HNTG	2 ml
	EDTA	0.1 ml
	Na <sub>3</sub> VO <sub>4</sub>	0.1 ml
25	Na <sub>4</sub> (P <sub>2</sub> O <sub>7</sub> )	0.1 ml
	H <sub>2</sub> O	7.3 ml

4. After drug incubation for two hours, transfer 10  $\mu$ l/well of 200nM IGF-1 Ligand in PBS to the cells (Final Conc. = 20 nM), and incubate at 5% CO<sub>2</sub> at 37 °C for 10 minutes.

5. Remove media and add 100  $\mu$ l/well HNTG\* and shake for 10 minutes. Look at cells under microscope to see if they are adequately lysed.
6. Use a 12-channel pipette to scrape the cells from the plate, and homogenize the lysate by repeated aspiration and dispensing. Transfer all the lysate to the antibody coated ELISA plate, and shake for 1 hour.
7. Remove the lysate, wash the plate, transfer anti-pTyr (1:3,000 with TBST) 100  $\mu$ l/well, and shake for 30 minutes.
8. Remove anti-pTyr, wash the plate, transfer TAGO (1:3,000 with TBST) 100  $\mu$ l/well, and shake for 30 minutes.
9. Remove detection antibody, wash the plate, and transfer fresh ABTS/H<sub>2</sub>O<sub>2</sub> (1.2  $\mu$ l H<sub>2</sub>O<sub>2</sub> to 10 ml ABTS) 100  $\mu$ l/well to the plate to start color development.
10. Measure OD at 410 nm with a reference wavelength of 630 nm in Dynatec MR5000.

#### 15 **EXAMPLE 15: EGF Receptor ELISA**

EGF Receptor kinase activity in cells genetically engineered to express human EGF-R was measured as described below:

#### 20 **Materials and Reagents**

The following materials and reagents were used:

1. EGF Ligand: stock concentration = 16.5  $\mu$ M; EGF 201, TOYOBO, Co., Ltd. Japan.
2. 05-101 (UBI) (a monoclonal antibody recognizing an EGFR extracellular domain).
- 25 3. Anti-phosphotyrosine antibody (anti-Ptyr) (polyclonal).
4. Detection antibody: Goat anti-rabbit IgG horse radish peroxidase conjugate, TAGO, Inc., Burlingame, CA.
5. TBST buffer:

30	Tris-HCl, pH 7	50 mM
	NaCl	150 mM

- |    |  |                                  |           |
|----|--|----------------------------------|-----------|
|    |  | Triton X-100                     | 0.1       |
| 6. |  | HNTG 5X stock:                   |           |
|    |  | HEPES                            | 0.1 M     |
|    |  | NaCl                             | 0.75 M    |
| 5  |  | Glycerol                         | 50        |
|    |  | Triton X-100                     | 1.0%      |
| 7. |  | ABTS stock:                      |           |
|    |  | Citric Acid                      | 100 mM    |
|    |  | Na <sub>2</sub> HPO <sub>4</sub> | 250 mM    |
| 10 |  | HCl, conc.                       | 4.0 pH    |
|    |  | ABTS*                            | 0.5 mg/ml |
- Keep solution in dark at 4 °C until used.
8. Stock reagents of:
- |    |  |  |
|----|--|--|
|    |  | EDTA 100 mM pH 7.0                                     |
| 15 |  | Na <sub>3</sub> VO <sub>4</sub> 0.5 M                  |
|    |  | Na <sub>4</sub> (P <sub>2</sub> O <sub>7</sub> ) 0.2 M |

### Protocol

The following protocol was used:

#### 20 A. Pre-coat ELISA Plate

1. Coat ELISA plates (Corning, 96 well, Cat. #25805-96) with 05-101 antibody at 0.5 µg per well in PBS, 150 µl final volume/well, and store overnight at 4 °C. Coated plates are good for up to 10 days when stored at 4 °C.
2. On day of use, remove coating buffer and replace with blocking buffer (5% Carnation Instant Non-Fat Dry Milk in PBS). Incubate the plate, shaking, at room temperature (about 23 °C to 25 °C) for 30 minutes. Just prior to use, remove blocking buffer and wash plate 4 times with TBST buffer.

#### B. Seeding Cells

1. NIH 3T3/C7 cell line (Honegger, *et al.*, 1987, *Cell* 51:199-209) can be use for this  
30 assay.

2. Choose dishes having 80-90% confluence for the experiment. Trypsinize cells and stop reaction by adding 10% CS DMEM medium. Suspend cells in DMEM medium (10% CS DMEM medium) and centrifuge once at 1000 rpm at room temperature for 5 minutes.

3. Resuspend cells in seeding medium (DMEM, 0.5% bovine serum), and count the cells using trypan blue. Viability above 90% is acceptable. Seed cells in DMEM medium (0.5% bovine serum) at a density of 10,000 cells per well, 100  $\mu$ l per well, in a 96 well microtiter plate. Incubate seeded cells in 5% CO<sub>2</sub> at 37 °C for about 40 hours.

### C. Assay Procedures.

1. Check seeded cells for contamination using an inverted microscope. Dilute drug stock (10 mg/ml in DMSO) 1:10 in DMEM medium, then transfer 5  $\mu$ l to a test well for a final drug dilution of 1:200 and a final DMSO concentration of 1%. Control wells receive DMSO alone. Incubate in 5% CO<sub>2</sub> at 37 °C for one hour.

2. Prepare EGF ligand: dilute stock EGF in DMEM so that upon transfer of 10  $\mu$ l dilute EGF (1:12 dilution), 25 nM final concentration is attained.

3. Prepare fresh 10 ml HNTG\* sufficient for 100  $\mu$ l per well wherein HNTG\* comprises: HNTG stock (2.0 ml), milli-Q H<sub>2</sub>O (7.3 ml), EDTA, 100 mM, pH 7.0 (0.5 ml), Na<sub>3</sub>VO<sub>4</sub> 0.5 M (0.1 ml) and Na<sub>4</sub>(P<sub>2</sub>O<sub>7</sub>), 0.2 M (0.1 ml).

4. Place on ice.

5. After two hours incubation with drug, add prepared EGF ligand to cells, 10  $\mu$ l per well, to yield a final concentration of 25 nM. Control wells receive DMEM alone. Incubate, shaking, at room temperature, for 5 minutes.

6. Remove drug, EGF, and DMEM. Wash cells twice with PBS. Transfer HNTG\* to cells, 100  $\mu$ l per well. Place on ice for 5 minutes. Meanwhile, remove blocking buffer from other ELISA plate and wash with TBST as described above.

7. With a pipette tip securely fitted to a micropipettor, scrape cells from plate and homogenize cell material by repeatedly aspirating and dispensing the HNTG\* lysis buffer. Transfer lysate to a coated, blocked, and washed ELISA plate. Incubate shaking at room temperature for one hour.

8. Remove lysate and wash 4 times with TBST. Transfer freshly diluted anti-Ptyr antibody to ELISA plate at 100  $\mu$ l per well. Incubate shaking at room temperature for 30 minutes in the presence of the anti-Ptyr antiserum (1:3000 dilution in TBST).
9. Remove the anti-Ptyr antibody and wash 4 times with TBST. Transfer the freshly  
5 diluted TAGO 30 anti-rabbit IgG antibody to the ELISA plate at 100  $\mu$ l per well. Incubate shaking at room temperature for 30 minutes (anti-rabbit IgG antibody: 1:3000 dilution in TBST).
10. Remove detection antibody and wash 4 times with TBST. Transfer freshly prepared ABTS/H<sub>2</sub>O<sub>2</sub> solution to ELISA plate, 100  $\mu$ l per well. Incubate at room temperature  
10 for 20 minutes. ABTS/H<sub>2</sub>O<sub>2</sub> solution: 1.2  $\mu$ l 30% H<sub>2</sub>O<sub>2</sub> in 10 ml ABTS stock.
11. Stop reaction by adding 50  $\mu$ l 5 N H<sub>2</sub>SO<sub>4</sub> (optional), and determine O.D. at 410 nm.
12. The maximal phosphotyrosine signal is determined by subtracting the value of the negative controls from the positive controls. The percent inhibition of phosphotyrosine content  
15 for extract-containing wells is then calculated, after subtraction of the negative controls.

#### **EXAMPLE 16: Met Autophosphorylation Assay – ELISA**

This assay determines Met tyrosine kinase activity by analyzing Met protein tyrosine  
20 kinase levels on the Met receptor.

##### **Materials and Reagents**

The following materials and reagents were used:

1. HNTG (5X stock solution): Dissolve 23.83 g HEPES and 43.83 g NaCl in about  
25 350 ml dH<sub>2</sub>O. Adjust pH to 7.2 with HCl or NaOH, add 500 ml glycerol and 10 ml Triton X-100, mix, add dH<sub>2</sub>O to 1 L total volume. To make 1 L of 1X working solution add 200 ml 5X stock solution to 800 ml dH<sub>2</sub>O, check and adjust pH as necessary, store at 4 °C.
2. PBS (Dulbecco's Phosphate-Buffered Saline), Gibco Cat. # 450-1300EB (1X solution).
- 30 3. Blocking Buffer: in 500 ml dH<sub>2</sub>O place 100 g BSA, 12.1 g Tris-pH7.5, 58.44 g NaCl and 10 ml Tween-20, dilute to 1 L total volume.

4. Kinase Buffer: To 500 ml dH<sub>2</sub>O add 12.1 g TRIS pH7.2, 58.4 g NaCl, 40.7 g MgCl<sub>2</sub> and 1.9 g EGTA; bring to 1 L total volume with dH<sub>2</sub>O.
5. PMSF (Phenylmethylsulfonyl fluoride), Sigma Cat. # P-7626, to 435.5 mg, add 100% ethanol to 25 ml total volume, vortex.
- 5 6. ATP (Bacterial Source), Sigma Cat. # A-7699, store powder at -20°C; to make up solution for use, dissolve 3.31 mg in 1 ml dH<sub>2</sub>O.
7. RC-20H HRPO Conjugated Anti-Phosphotyrosine, Transduction Laboratories Cat. # E120H.
8. Pierce 1-Step (TM) Turbo TMB-ELISA (3,3',5,5'-tetramethylbenzidine, Pierce  
10 Cat. # 34022.
9. H<sub>2</sub>SO<sub>4</sub>, add 1 ml conc. (18 N) to 35 ml dH<sub>2</sub>O.
10. Tris-HCl, Fischer Cat. # BP152-5; to 121.14 g of material, add 600 ml MilliQ H<sub>2</sub>O, adjust pH to 7.5 (or 7.2) with HCl, bring volume to 1 L with MilliQ H<sub>2</sub>O.
11. NaCl, Fischer Cat. # S271-10, make up 5 M solution.
- 15 12. Tween-20, Fischer Cat. # S337 -500.
13. Na<sub>3</sub>VO<sub>4</sub>, Fischer Cat. # S454-50, to 1.8 g material add 80 ml MilliQ H<sub>2</sub>O, adjust pH to 10.0 with HCl or NaOH, boil in microwave, cool, check pH, repeat procedure until pH stable at 10.0, add MilliQ H<sub>2</sub>O to 100 ml total volume, make 1 ml aliquots and store at -80°C.
14. MgCl<sub>2</sub>, Fischer Cat. # M33-500, make up 1 M solution.
- 20 15. HEPES, Fischer Cat. # BP310-500, to 200 ml MilliQ H<sub>2</sub>O, add 59.6 g material, adjust pH to 7.5, bring volume to 250 ml total, sterile filter.
16. Albumin, Bovine (BSA), Sigma Cat. # A-4503, to 30 grams material add sterile distilled water to make total volume of 300 ml, store at 4 °C.
17. TBST Buffer: to approx. 900 ml dH<sub>2</sub>O in a 1 L graduated cylinder add 6.057 g  
25 TRIS and 8.766 g NaCl, when dissolved, adjust pH to 7.2 with HCl, add 1.0 ml Triton X-100 and bring to 1 L total volume with dH<sub>2</sub>O.
18. Goat Affinity purified antibody Rabbit IgG (whole molecule), Cappel Cat. # 55641.
19. Anti h-Met (C-28) rabbit polyclonal IgG antibody, Santa Cruz Chemical Cat. #  
30 SC-161.

20. Transiently Transfected EGFR/Met chimeric cells (EMR) (Komada, *et al.*, 1993, *Oncogene* 8:2381-2390.

21. Sodium Carbonate Buffer, ( $\text{Na}_2\text{CO}_3$ , Fischer Cat. # S495): to 10.6 g material add 800 ml MilliQ  $\text{H}_2\text{O}$ , when dissolved adjust pH to 9.6 with NaOH, bring up to 1 L total volume with MilliQ  $\text{H}_2\text{O}$ , filter, store at 4 °C.

### Procedure

All of the following steps are conducted at room temperature unless it is specifically indicated otherwise. All ELISA plate washing is by rinsing 4X with TBST.

#### 10 A. EMR Lysis

This procedure can be performed the night before or immediately prior to the start of receptor capture.

1. Quick thaw lysates in a 37 °C waterbath with a swirling motion until the last crystals disappear.
- 15 2. Lyse cell pellet with 1X HNTG containing 1 mM PMSF. Use 3 ml of HNTG per 15 cm dish of cells. Add ½ the calculated HNTG volume, vortex the tube for 1 min., add the remaining amount of HNTG, vortex for another min.
3. Balance tubes, centrifuge at 10,000x g for 10 min at 4 °C.
4. Pool supernatants, remove an aliquot for protein determination.
- 20 5. Quick freeze pooled sample in dry ice/ethanol bath. This step is performed regardless of whether lysate will be stored overnight or used immediately following protein determination.
6. Perform protein determination using standard bicinchoninic acid (BCA) method (BCA Assay Reagent Kit from Pierce Chemical Cat. # 23225).

#### 25 B. ELISA Procedure

1. Coat Corning 96 well ELISA plates with 5 µg per well Goat anti-Rabbit antibody in Carbonate Buffer for a total well volume of 50 µl. Store overnight at 4 °C.
2. Remove unbound Goat anti-rabbit antibody by inverting plate to remove liquid.
3. Add 150 µl of Blocking Buffer to each well. Incubate for 30 min. at room
- 30 temperature with shaking.



4. Wash 4X with TBST. Pat plate on a paper towel to remove excess liquid and bubbles.
5. Add 1  $\mu\text{g}$  per well of Rabbit anti-Met antibody diluted in TBST for a total well volume of 100  $\mu\text{l}$ .
- 5 6. Dilute lysate in HNTG (90  $\mu\text{g}$  lysate/100  $\mu\text{l}$ )
7. Add 100  $\mu\text{l}$  of diluted lysate to each well. Shake at room temperature for 60 min.
8. Wash 4X with TBST. Pat on paper towel to remove excess liquid and bubbles.
9. Add 50  $\mu\text{l}$  of 1X lysate buffer per well.
- 10 10. Dilute compounds/extracts 1:10 in 1X Kinase Buffer in a polypropylene 96 well plate.
11. Transfer 5.5  $\mu\text{l}$  of diluted drug to ELISA plate wells. Incubate at room temperature with shaking for 20 min.
12. Add 5.5  $\mu\text{l}$  of 60  $\mu\text{M}$  ATP solution per well. Negative controls do not receive any ATP. Incubate at room temperature for 90 min., with shaking.
- 15 13. Wash 4X with TBST. Pat plate on paper towel to remove excess liquid and bubbles.
14. Add 100  $\mu\text{l}$  per well of RC20 (1:3000 dilution in Blocking Buffer). Incubate 30 min. at room temperature with shaking.
15. Wash 4X with TBST. Pat plate on paper towel to remove excess liquid and bubbles.
- 20 16. Add 100  $\mu\text{l}$  per well of Turbo-TMB. Incubate with shaking for 30-60 min.
17. Add 100  $\mu\text{l}$  per well of 1 M  $\text{H}_2\text{SO}_4$  to stop reaction.
18. Read assay on Dynatech MR7000 ELISA reader. Test Filter = 450 nm, reference filter = 410 nm.

25

**EXAMPLE 17: Biochemical *src* Assay – ELISA**

This assay is used to determine *src* protein kinase activity measuring phosphorylation of a biotinylated peptide as the readout.

30

**Materials and Reagents**

The following materials and reagents were used:

1. Yeast transformed with *src*.
2. Cell lysates: Yeast cells expressing *src* are pelleted, washed once with water, re-pelleted and stored at -80°C until use.
- 5 3. N-terminus biotinylated EEEYEEYEEYEEYEEYEEY is prepared by standard procedures well known to those skilled in the art.
4. DMSO: Sigma, St. Louis, MO.
5. 96 Well ELISA Plate: Corning 96 Well Easy Wash, Modified flat Bottom Plate, Corning Cat. #25805-96.
- 10 6. NUNC 96-well V-bottom polypropylene plates for dilution of compounds: Applied Scientific Cat. # A-72092.
7. Vecastain ELITE ABC reagent: Vector, Burlingame, CA.
8. Anti-*src* (327) mab: Schizosaccharomyces Pombe was used to express recombinant *src* (Superti-Furga, *et al.*, *EMBO J.* 12:2625-2634; Superti-Furga, *et al.*, *Nature*  
15 *Biochem.* 14:600-605). S. Pombe strain SP200 (h-s leul.32 ura4 ade210) was grown as described and transformations were pRSP expression plasmids were done by the lithium acetate method (Superti-Furga, *supra*). Cells were grown in the presence of 1 µM thiamin to repress expression from the nm1 promoter or in the absence of thiamin to induce expression.
9. Monoclonal anti-phosphotyrosine, UBI 05-321 (UB40 may be used instead).
- 20 10. Turbo TMB-ELISA peroxidase substrate: Pierce Chemical.

Buffer Solutions:

1. PBS (Dulbecco's Phosphate-Buffered Saline): GIBCO PBS, GIBCO Cat. # 450-1300EB.
- 25 2. Blocking Buffer: 5% Non-fat milk (Carnation) in PBS.
3. Carbonate Buffer: Na<sub>2</sub>CO<sub>3</sub> from Fischer, Cat. # S495, make up 100 mM stock solution.
4. Kinase Buffer: 1.0 ml (from 1 M stock solution) MgCl<sub>2</sub>; 0.2 ml (from a 1 M stock solution) MnCl<sub>2</sub>; 0.2 ml (from a 1 M stock solution) DTT; 5.0 ml (from a 1 M stock  
30 solution) HEPES; 0.1 ml TX-100; bring to 10 ml total volume with MilliQ H<sub>2</sub>O.

5. Lysis Buffer: 5.0 HEPES (from 1 M stock solution.); 2.74 ml NaCl (from 5 M stock solution); 10 ml glycerol; 1.0 ml TX-100; 0.4 ml EDTA (from a 100 mM stock solution); 1.0 ml PMSF (from a 100 mM stock solution); 0.1 ml  $\text{Na}_3\text{VO}_4$  (from a 0.1 M stock solution); bring to 100 ml total volume with MilliQ  $\text{H}_2\text{O}$ .
- 5 6. ATP: Sigma Cat. # A-7699, make up 10 mM stock solution (5.51 mg/ml).
7. TRIS-HCl: Fischer Cat. # BP 152-5, to 600 ml MilliQ  $\text{H}_2\text{O}$  add 121.14 g material, adjust pH to 7.5 with HCl, bring to 1 L total volume with MilliQ  $\text{H}_2\text{O}$ .
8. NaCl: Fischer Cat. # S271-10, Make up 5 M stock solution with MilliQ  $\text{H}_2\text{O}$ .
9.  $\text{Na}_3\text{VO}_4$ : Fischer Cat. # S454-50; to 80 ml MilliQ  $\text{H}_2\text{O}$ , add 1.8 g material; adjust  
10 pH to 10.0 with HCl or NaOH; boil in a microwave; cool; check pH, repeat pH adjustment until pH remains stable after heating/cooling cycle; bring to 100 ml total volume with MilliQ  $\text{H}_2\text{O}$ ; make 1 ml aliquots and store at  $-80^\circ\text{C}$ .
10.  $\text{MgCl}_2$ : Fischer Cat. # M33-500, make up 1 M stock solution with MilliQ  $\text{H}_2\text{O}$ .
11. HEPES: Fischer Cat. # BP 310-500; to 200 ml MilliQ  $\text{H}_2\text{O}$ , add 59.6 g material,  
15 adjust pH to 7.5, bring to 250 ml total volume with MilliQ  $\text{H}_2\text{O}$ , sterile filter (1 M stock solution).
12. TBST Buffer: TBST Buffer: To 900 ml  $\text{dH}_2\text{O}$  add 6.057 g TRIS and 8.766 g NaCl; adjust pH to 7.2 with HCl, add 1.0 ml Triton-X-100; bring to 1 L total volume with  $\text{dH}_2\text{O}$ .
13.  $\text{MnCl}_2$ : Fischer Cat. # M87-100, make up 1 M stock solution with MilliQ  $\text{H}_2\text{O}$ .
- 20 14. DTT; Fischer Cat. # BP172-5.
15. TBS (TRIS Buffered Saline): to 900 ml MilliQ  $\text{H}_2\text{O}$  add 6.057 g TRIS and 8.777 g NaCl; bring to 1 L total volume with MilliQ  $\text{H}_2\text{O}$ .
16. Kinase Reaction Mixture: Amount per assay plate (100 wells): 1.0 ml Kinase Buffer, 200  $\mu\text{g}$  GST- $\zeta$ , bring to final volume of 8.0 ml with MilliQ  $\text{H}_2\text{O}$ .
- 25 17. Biotin labeled EEEYEEYEEYEEYEEY: Make peptide stock solution (1 mM, 2.98 mg/ml) in water fresh just before use.
18. Vectastain ELITE ABC reagent: To prepare 14 ml of working reagent, add 1 drop of reagent A to 15 ml TBST and invert tube several times to mix. Then add 1 drop of reagent B. Put tube on orbital shaker at room temperature and mix for 30 minutes.

30

ProtocolA. Preparation of *src* coated ELISA plate.

1. Coat ELISA plate with 0.5  $\mu\text{g}$ /well anti-*src* mab in 100  $\mu\text{l}$  of pH 9.6 sodium carbonate buffer at 4  $^{\circ}\text{C}$  overnight.
- 5      2. Wash wells once with PBS.
3. Block plate with 0.15 ml 5% milk in PBS for 30 min. at room temperature.
4. Wash plate 5X with PBS.
5. Add 10  $\mu\text{g}$ /well of *src* transformed yeast lysates diluted in Lysis Buffer (0.1 ml total volume per well). (Amount of lysate may vary between batches.) Shake plate for 20  
10 minutes at room temperature.

B. Preparation of phosphotyrosine antibody-coated ELISA plate.

1. 4G10 plate: coat 0.5  $\mu\text{g}$ /well 4G10 in 100  $\mu\text{l}$  PBS overnight at 4  $^{\circ}\text{C}$  and block with 150  $\mu\text{l}$  of 5% milk in PBS for 30 minutes at room temperature.

C. Kinase assay procedure.

- 15      1. Remove unbound proteins from step 1-7, above, and wash plates 5X with PBS.
2. Add 0.08 ml Kinase Reaction Mixture per well (containing 10  $\mu\text{l}$  of 10X Kinase Buffer and 10  $\mu\text{M}$  (final concentration) biotin-EEEEYEEYEEEEYEEY per well diluted in water.
3. Add 10  $\mu\text{l}$  of compound diluted in water containing 10% DMSO and pre-incubate  
20 for 15 minutes at room temperature.
4. Start kinase reaction by adding 10  $\mu\text{l}$ /well of 0.05 mM ATP in water (5  $\mu\text{M}$  ATP final).
5. Shake ELISA plate for 15 min. at room temperature.
6. Stop kinase reaction by adding 10  $\mu\text{l}$  of 0.5 M EDTA per well.
- 25      7. Transfer 90  $\mu\text{l}$  supernatant to a blocked 4G10 coated ELISA plate from section B, above.
8. Incubate for 30 min. while shaking at room temperature.
9. Wash plate 5X with TBST.

10. Incubate with Vectastain ELITE ABC reagent (100 µl/well) for 30 min. at room temperature.
11. Wash the wells 5X with TBST.
12. Develop with Turbo TMB.

5

**EXAMPLE 18: Biochemical *lck* Assay – ELISA**

This assay is used to determine *lck* protein kinase activities measuring phosphorylation of GST- $\zeta$  as the readout.

10

**Materials and Reagents**

The following materials and reagents were used:

1. Yeast transformed with *lck*. *Schizosaccharomyces Pombe* was used to express recombinant *lck* (Superti-Furga, *et al.*, *EMBO J.* 12:2625-2634; Superti-Furga, *et al.*, *Nature Biotech.* 14:600-605). *S. Pombe* strain SP200 (h-s leul.32 ura4 ade210) was grown as described and transformations with pRSP expression plasmids were done by the lithium acetate method (Superti-Furga, *supra*). Cells were grown in the presence of 1 µM thiamin to induce expression.
2. Cell lysates: Yeast cells expressing *lck* are pelleted, washed once in water, re-pelleted and stored frozen at -80°C until use.
- 20 3. GST- $\zeta$ : DNA encoding for GST- $\zeta$  fusion protein for expression in bacteria obtained from Arthur Weiss of the Howard Hughes Medical Institute at the University of California, San Francisco. Transformed bacteria were grown overnight while shaking at 25°C. GST- $\zeta$  was purified by glutathione affinity chromatography, Pharmacia, Alameda, CA.
4. DMSO: Sigma, St. Louis, MO.
- 25 5. 96-Well ELISA plate: Corning 96 Well Easy Wash, Modified Flat Bottom Plate, Corning Cat. #25805-96.
6. NUNC 96-well V-bottom polypropylene plates for dilution of compounds: Applied Scientific Cat. # AS-72092.
7. Purified Rabbit anti-GST antiserum: Amrad Corporation (Australia) Cat.
- 30 #90001605.
8. Goat anti-Rabbit-IgG-HRP: Amersham Cat. # V010301

9. Sheep ant-mouse IgG (H+L): Jackson Labs Cat. # 5215-005-003.
10. Anti-*lck* (3A5) mab: Santa Cruz Biotechnology Cat # sc-433.
11. Monoclonal anti-phosphotyrosine UBI 05-321 (UB40 may be used instead).

5      Buffer solutions:

1. PBS (Dulbecco's Phosphate-Buffered Saline) 1X solution: GIBCO PBS, GIBCO Cat. # 450-1300EB.
2. Blocking Buffer: 100 g BSA, 12.1 g. TRIS-pH7.5, 58.44 g NaCl, 10 ml Tween-20, bring up to 1 L total volume with MilliQ H<sub>2</sub>O.
- 10      3. Carbonate Buffer: Na<sub>2</sub>CO<sub>4</sub> from Fischer, Cat. # S495; make up 100 mM solution with MilliQ H<sub>2</sub>O.
4. Kinase Buffer: 1.0 ml (from 1 M stock solution) MgCl<sub>2</sub>; 0.2 ml (from a 1 M stock solution) MnCl<sub>2</sub>; 0.2 ml (from a 1 M stock solution) DTT; 5.0 ml (from a 1 M stock solution) HEPES; 0.1 ml TX-100; bring to 10 ml total volume with MilliQ H<sub>2</sub>O.
- 15      5. Lysis Buffer: 5.0 HEPES (from 1 M stock solution.); 2.74 ml NaCl (from 5 M stock solution); 10 ml glycerol; 1.0 ml TX-100; 0.4 ml EDTA (from a 100 mM stock solution); 1.0 ml PMSF (from a 100 mM stock solution); 0.1 ml Na<sub>3</sub>VO<sub>4</sub> (from a 0.1 M stock solution); bring to 100 ml total volume with MilliQ H<sub>2</sub>O.
6. ATP: Sigma Cat. # A-7699, make up 10 mM stock solution (5.51 mg/ml).
- 20      7. TRIS-HCl: Fischer Cat. # BP 152-5, to 600 ml MilliQ H<sub>2</sub>O add 121.14 g material, adjust pH to 7.5 with HCl, bring to 1 L total volume with MilliQ H<sub>2</sub>O.
8. NaCl: Fischer Cat. # S271-10, Make up 5 M stock solution with MilliQ H<sub>2</sub>O.
9. Na<sub>3</sub>VO<sub>4</sub>: Fischer Cat. # S454-50; to 80 ml MilliQ H<sub>2</sub>O, add 1.8 g material; adjust pH to 10.0 with HCl or NaOH; boil in a microwave; cool; check pH, repeat pH adjustment until
- 25      pH remains stable after heating/cooling cycle; bring to 100 ml total volume with MilliQ H<sub>2</sub>O; make 1 ml aliquots and store at -80°C.
10. MgCl<sub>2</sub>: Fischer Cat. # M33-500, make up 1 M stock solution with MilliQ H<sub>2</sub>O.
11. HEPES: Fischer Cat. # BP 310-500; to 200 ml MilliQ H<sub>2</sub>O, add 59.6 g material, adjust pH to 7.5, bring to 250 ml total volume with MilliQ H<sub>2</sub>O, sterile filter (1M stock solution).

12. Albumin, Bovine (BSA), Sigma Cat. # A4503; to 150 ml MilliQ H<sub>2</sub>O add 30 g material, bring 300 ml total volume with MilliQ H<sub>2</sub>O, filter through 0.22  $\mu$ m filter, store at 4 °C.
13. TBST Buffer: To 900 ml dH<sub>2</sub>O add 6.057 g TRIS and 8.766 g NaCl; adjust pH to 7.2 with HCl, add 1.0 ml Triton-X-100; bring to 1 L total volume with dH<sub>2</sub>O.
- 5 14. MnCl<sub>2</sub>: Fischer Cat. # M87-100, make up 1 M stock solution with MilliQ H<sub>2</sub>O.
15. DTT; Fischer Cat. # BP172-5.
16. TBS (TRIS Buffered Saline): to 900 ml MilliQ H<sub>2</sub>O add 6.057 g TRIS and 8.777 g NaCl; bring to 1 L total volume with MilliQ H<sub>2</sub>O.
17. Kinase Reaction Mixture: Amount per assay plate (100 wells): 1.0 ml Kinase
- 10 Buffer, 200  $\mu$ g GST- $\zeta$ , bring to final volume of 8.0 ml with MilliQ H<sub>2</sub>O.

#### Procedures

##### A. Preparation of *lck* coated ELISA plate.

1. Coat 2.0  $\mu$ g/well Sheep anti-mouse IgG in 100  $\mu$ l of pH 9.6 sodium carbonate
- 15 buffer at 4 °C overnight.
2. Wash well once with PBS.
3. Block plate with 0.15 ml of blocking Buffer for 30 min. at room temp.
4. Wash plate 5X with PBS.
5. Add 0.5  $\mu$ g/well of anti-*lck* (mab 3A5) in 0.1 ml PBS at room temperature for 1-2
- 20 hours.
6. Wash plate 5X with PBS.
7. Add 20  $\mu$ g/well of *lck* transformed yeast lysates diluted in Lysis Buffer (0.1 ml total volume per well). (Amount of lysate may vary between batches) Shake plate at 4 °C overnight to prevent loss of activity.

##### B. Preparation of phosphotyrosine antibody-coated ELISA plate.

1. UB40 plate: 1.0  $\mu$ g/well UB40 in 100  $\mu$ l of PBS overnight at 4 °C and block with 150  $\mu$ l of Blocking Buffer for at least 1 hour.

##### C. Kinase assay procedure.

1. Remove unbound proteins from step 1-7, above, and wash plates 5X with PBS.

2. Add 0.08 ml Kinase Reaction Mixture per well (containing 10  $\mu$ l of 10X Kinase Buffer and 2  $\mu$ g GST- $\zeta$  per well diluted with water).
3. Add 10  $\mu$ l of compound diluted in water containing 10% DMSO and pre-incubate for 15 minutes at room temperature.
- 5 4. Start kinase reaction by adding 10  $\mu$ l/well of 0.1 mM ATP in water (10  $\mu$ M ATP final).
5. Shake ELISA plate for 60 min. at room temperature.
6. Stop kinase reaction by adding 10  $\mu$ l of 0.5 M EDTA per well.
7. Transfer 90  $\mu$ l supernatant to a blocked 4G10 coated ELISA plate from section B,
- 10 above.
8. Incubate while shaking for 30 min. at room temperature.
9. Wash plate 5X with TBST.
10. Incubate with Rabbit anti-GST antibody at 1:5000 dilution in 100  $\mu$ l TBST for 30 min. at room temperature.
- 15 11. Wash the wells 5X with TBST.
12. Incubate with Goat anti-Rabbit-IgG-HRP at 1:20,000 dilution in 100  $\mu$ l of TBST for 30 min. at room temperature.
13. Wash the wells 5X with TBST.
14. Develop with Turbo TMB.

20

#### **EXAMPLE 19: Biochemical c-kit Assay – ELISA**

##### **A. Materials And Reagents**

- 1) HNTG: 5X stock concentration: 100 mM HEPES pH 7.2, 750 mM NaCl, 50%  
25 glycerol, 2.5% Triton X-100.
- 2) PBS (Dulbecco's Phosphate-Buffered Saline): Gibco Catalog # 450-1300EB
- 3) 1 X Blocking Buffer: 10 mM TRIS-pH7.5, 1 % BSA, 100 mM NaCl, 0.1%  
Triton X-100
- 4) 1 X Kinase Buffer: 25 mM HEPES, 100 mM NaCl, 10 mM Mg Cl<sub>2</sub>, 6 mM Mn  
30 Cl<sub>2</sub>.



- 5) PMSF: Stock Solution = 100mM (Sigma Catalog # P-7626)
- 6) 10 mM ATP (Bacterial source) Sigma A-7699, 5g.
- 7) UB40 anti-phosphotyrosine mAb (available from Terrance at Sugen.
- 8) HRP conjugated sheep anti-Mouse IgG. (Amersham NA 931)
- 5 9) ABTS (5Prime-3Prime 7-579844)
- 10) TRIS HCL: Fisher BP 152-5
- 11) NaCl: Fisher S271-10
- 12) Triton X-100: Fisher BP151-100
- 13) Na<sub>3</sub>VO<sub>4</sub>: Fisher S454-50
- 10 14) MgCl<sub>2</sub>: Fisher M33-500
- 15) MnCl<sub>2</sub>: Fisher M87-500
- 16) HEPES: Fisher BP310-500
- 17) Albumin, Bovine (BSA): Sigma A-8551
- 18) TBST Buffer: 50 mM Tris pH 7.2, 150 mM NaCl, 0.1% Triton X-100.
- 15 19) Goat affinity purified antibody Rabbit IgG (whole molecule): Cappel 55641.
- 20) Anti Kit (C-20) rabbit polyclonal IgG antibody: Santa Cruz sc-168
- 21) Kit/CHO cells: CHO cells stably expressing GyrB/Kit, which are grown in standard CHO medium, supplemented with 1mg/ml G418
- 22) Indolinone Compounds: The indolinone compounds were synthesized as set forth in the following application: PCT application number US99/06468, filed March 26, 1999 by Fong, *et al.* and entitled METHODS OF MODULATING TYROSINE PROTEIN KINASE (Lyon & Lyon docket number 231/250 PCT which is hereby incorporated by reference in its entirety including any drawings.

#### B. Procedure

- 25 All of the following steps are conducted at room temperature unless it is specifically indicated. All ELISA plate washing is by rinsing 4x with TBST.

#### Kit Cell Lysis

This procedure is performed 1hour prior to the start of receptor capture.

- 30 1) Wash a >95% confluent 15 cm dish with PBS and aspirate as much as possible.

- 2) Lyse the cells with 3 ml of 1x HNTG containing 1 mM PMSF/15 cm dish. Scrape the cells from the plate and transfer to a 50 ml centrifuge tube.
- 3) Pool supernatants, and allow to sit, on ice, for one hour with occasional vortexing. Failure to do so with result in an increased background (approximately 3-fold higher).
- 5 4) Balance tubes and centrifuge at 10,000 x g for 10 min at 4 °C. Remove an aliquot for protein determination
- 5) Perform protein determination as per the SOP for protein determination using the bicinchoninic acid (BCA) method.

#### 10 ELISA Procedure

- 1) Coat Corning 96-well ELISA plates with 2 µg per well Goat anti-rabbit antibody in PBS for a total well volume of 100 µl. Store overnight at 4 °C.
- 2) Remove unbound Goat anti-rabbit antibody by inverting plate to remove liquid.
- 3) Add 100 µl of Blocking Buffer to each well. Shake at room temperature for 60  
15 min.
- 4) Wash 4x with TBST. Pat plate on a paper towel to remove excess liquid and bubbles
- 5) Add 0.2 µg per well of Rabbit anti -Kit antibody diluted in TBST for a total well volume of 100 µl. Shake at room temperature for 60 min.
- 20 6) Dilute lysate in HNTG (180 µg lysate/100 µl)
- 7) Add 100 µl of diluted lysate to each well. Shake at room temperature for 60 min.
- 8) Wash 4x with TBST. Pat plate on a paper towel to remove excess liquid and bubbles.
- 9) Dilute compounds/extracts (or as stated otherwise) in 1x kinase buffer, with 5µM  
25 ATP in a polypropylene 96 well plate.
- 10) Transfer 100 µl of diluted drug to ELISA plate wells. Incubate at room temperature with shaking for 60 min.
- 11) Stop reaction with the addition of 10 µl of 0.5 M EDTA. Plate is now stable for a reasonable period of time.

- 12) Wash 4x with TBST. Pat plate on a paper towel to remove excess liquid and bubbles.
- 13) Add 100 µl per well of UB40 (1:2000 dilution in TBST). Incubate 60 min at room temperature, with shaking.
- 5 14) Wash 4x with TBST. Pat plate on a paper towel to remove excess liquid and bubbles.
- 15) Add 100 µl per well of sheep anti-mouse IgG - HRP (1:5000 dilution in TBST). Incubate 60 min at room temperature, with shaking.
- 16) Wash 4x with TBST. Pat plate on a paper towel to remove excess liquid and  
10 bubbles.
- 17) Add 100 µl per well of ABTS. Incubate with shaking for 15-30 min.
- 18) Read assay on Dynatech MR7000 ELISA reader
- Test Filter = 410 nm
- Reference Filter = 630 nm.

15

**EXAMPLE 20: Assay Measuring Phosphorylating Function of RAF**

The following assay reports the amount of RAF-catalyzed phosphorylation of its target protein MEK as well as MEK's target MAPK. The RAF gene sequence is described in Bonner  
20 *et al.*, 1985, *Molec. Cell. Biol.* 5:1400-1407, and is readily accessible in multiple gene sequence data banks. Construction of the nucleic acid vector and cell lines utilized for this portion of the invention are fully described in Morrison *et al.*, 1988, *Proc. Natl. Acad. Sci. USA* 85:8855-8859.

**Materials and Reagents**

- 25 1. *Sf9* (*Spodoptera frugiperda*) cells; GIBCO-BRL, Gaithersburg, MD.
2. RIPA buffer: 20 mM Tris/HCl pH 7.4, 137 mM NaCl, 10% glycerol, 1 mM PMSF, 5 mg/L Aprotinin, 0.5 % Triton X-100.
3. Thioredoxin-MEK fusion protein (T-MEK): T-MEK expression and purification by affinity chromatography were performed according to the manufacturer's procedures.
- 30 Catalog# K 350-01 and R 350-40, Invitrogen Corp., San Diego, CA.

4. His-MAPK (ERK 2); His-tagged MAPK was expressed in XL1 Blue cells transformed with pUC18 vector encoding His-MAPK. His-MAPK was purified by Ni-affinity chromatography. Cat# 27-4949-01, Pharmacia, Alameda, CA, as described herein.
5. Sheep anti mouse IgG: Jackson laboratories, West Grove, PA. Catalog, # 515-006-008, Lot# 28563.
6. RAF-1 protein kinase specific antibody: URP2653 from UBI.
7. Coating buffer: PBS; phosphate buffered saline, GIBCO-BRL, Gaithersburg, MD.
8. Wash buffer: TBST - 50 mM Tris/HCl pH 7.2, 150 mM NaCl, 0.1 % Triton X-100.
9. Block buffer: TBST, 0.1 % ethanolamine pH 7.4.
10. DMSO, Sigma, St. Louis, MO.
11. Kinase buffer (KB): 20 mM HEPES/HCl pH 7.2, 150 mM NaCl, 0.1 % Triton X-100, 1 mM PMSF, 5 mg/L Aprotinin, 75 mM sodium ortho vanadate, 0.5 mM DTT and 10 mM  $MgCl_2$ .
12. ATP mix: 100 mM  $MgCl_2$ , 300 mM ATP, 10 mCi 33P ATP (Dupont-NEN)/ml.
13. Stop solution: 1 % phosphoric acid; Fisher, Pittsburgh, PA.
14. Wallac Cellulose Phosphate Filter mats; Wallac, Turku, Finland.
15. Filter wash solution: 1 % phosphoric acid, Fisher, Pittsburgh, PA.
16. Tomtec plate harvester, Wallac, Turku, Finland.
17. Wallac beta plate reader # 1205, Wallac, Turku, Finland.
18. NUNC 96-well V bottom polypropylene plates for compounds Applied Scientific Catalog # AS-72092.

#### Protocol

- 25 All of the following steps were conducted at room temperature unless specifically indicated.

1. ELISA plate coating: ELISA wells are coated with 100 ml of Sheep anti mouse affinity purified antiserum (1 mg/100 ml coating buffer) over night at 4 °C. ELISA plates can be used for two weeks when stored at 4 °C.

2. Invert the plate and remove liquid. Add 100 ml of blocking solution and incubate for 30 min.
3. Remove blocking solution and wash four times with wash buffer. Pat the plate on a paper towel to remove excess liquid.
- 5 4. Add 1 mg of antibody specific for RAF-1 to each well and incubate for 1 hour. Wash as described in step 3.
- 10 5. Thaw lysates from RAS/RAF infected Sf9 cells and dilute with TBST to 10 mg/100 ml. Add 10 mg of diluted lysate to the wells and incubate for 1 hour. Shake the plate during incubation. Negative controls receive no lysate. Lysates from RAS/RAF infected Sf9 insect cells are prepared after cells are infected with recombinant baculoviruses at a MOI of 5 for each virus, and harvested 48 hours later. The cells are washed once with PBS and lysed in RIPA buffer. Insoluble material is removed by centrifugation (5 min at 10,000 x g). Aliquots of lysates are frozen in dry ice/ethanol and stored at -80°C until use.
- 15 6. Remove non-bound material and wash as outlined above (step 3).
7. Add 2 mg of T-MEK and 2 mg of His-MAEPK per well and adjust the volume to 40 ml with kinase buffer. Methods for purifying T-MEK and MAPK from cell extracts are provided herein by example.
- 20 8. Pre-dilute compounds (stock solution 10 mg/ml DMSO) or extracts 20 fold in TBST plus 1% DMSO. Add 5 ml of the pre-diluted compounds/extracts to the wells described in step 6. Incubate for 20 min. Controls receive no drug.
9. Start the kinase reaction by addition of 5 ml ATPmix; Shake the plates on an ELISA plate shaker during incubation.
10. Stop the kinase reaction after 60 min by addition of 30 ml stop solution to each well.
- 25 11. Place the phosphocellulose mat and the ELISA plate in the Tomtec plate harvester. Harvest and wash the filter with the filter wash solution according to the manufacturers recommendation. Dry the filter mats. Seal the filter mats and place them in the holder. Insert the holder into radioactive detection apparatus and quantify the radioactive phosphorous on the filter mats.

Alternatively, 40 ml aliquots from individual wells of the assay plate can be transferred to the corresponding positions on the phosphocellulose filter mat. After air drying the filters, put the filters in a tray. Gently rock the tray, changing the wash solution at 15 min intervals for 1 hour. Air-dry the filter mats. Seal the filter mats and place them in a holder suitable for measuring the radioactive phosphorous in the samples. Insert the holder into a detection device and quantify the radioactive phosphorous on the filter mats.

#### **EXAMPLE 21: CDK2/Cyclin A - Inhibition Assay**

10        This assay analyzes the protein kinase activity of CDK2 in exogenous substrate.

##### **Materials and Reagents**

1.        Buffer A (80 mM Tris ( pH 7.2), 40 mM  $MgCl_2$ ): 4.84 g Tris (F.W. =121.1 g/mol), 4.07 g  $MgCl_2$  (F.W.=203.31 g/mol) dissolved in 500 ml  $H_2O$ . Adjust pH to 7.2 with HCl.
2.        Histone H1 solution (0.45 mg/ml Histone H1 and 20 mM HEPES pH 7.2: 5 mg Histone H1 (Boehinger Mannheim) in 11.111 ml 20 mM HEPES pH 7.2 (477 mg HEPES (F.W.= 238.3 g/mol) dissolved in 100 ml dd $H_2O$ ), stored in 1 ml aliquots at -80°C.
3.        ATP solution (60  $\mu$ M ATP, 300  $\mu$ g/ml BSA, 3 mM DTT): 120  $\mu$ l 10 mM ATP, 600  $\mu$ l 10 mg/ml BSA to 20 ml, stored in 1 ml aliquots at -80°C.
4.        CDK2 solution: cdk2/cyclin A in 10 mM HEPES pH 7.2, 25 mM NaCl, 0.5 mM DTT, 10% glycerol, stored in 9  $\mu$ l aliquots at -80°C.

##### **Description of Assay:**

1.        Prepare solutions of inhibitors at three times the desired final assay concentration in dd $H_2O$ /15 % DMSO by volume.
2.        Dispense 20  $\mu$ l of inhibitors to wells of polypropylene 96-well plates (or 20  $\mu$ l 15% DMSO for positive and negative controls).
3.        Thaw Histone H1 solution (1 ml/plate), ATP solution (1 ml/plate plus 1 aliquot for negative control), and CDK2 solution (9  $\mu$ /plate). Keep CDK2 on ice until use. Aliquot CDK2 solution appropriately to avoid repeated freeze-thaw cycles.

4. Dilute 9  $\mu$ l CDK2 solution into 2.1 ml Buffer A (per plate). Mix. Dispense 20  $\mu$ l into each well.
5. Mix 1 ml Histone H1 solution with 1 ml ATP solution (per plate) into a 10 ml screw cap tube. Add  $\gamma^{33}\text{P}$  ATP to a concentration of 0.15  $\mu\text{Ci}/20 \mu\text{l}$  (0.15  $\mu\text{Ci}/\text{well}$  in assay).
- 5 Mix carefully to avoid BSA frothing. Add 20  $\mu$ l to appropriate wells. Mix plates on plate shaker. For negative control, mix ATP solution with an equal amount of 20 mM HEPES pH 7.2 and add  $\gamma^{33}\text{P}$  ATP to a concentration of 0.15  $\mu\text{Ci}/20 \mu\text{l}$  solution. Add 20  $\mu$ l to appropriate wells.
6. Let reactions proceed for 60 minutes.
7. Add 35  $\mu$ l 10% TCA to each well. Mix plates on plate shaker.
- 10 8. Spot 40  $\mu$ l of each sample onto P30 filter mat squares. Allow mats to dry (approx. 10-20 minutes).
9. Wash filter mats 4 X 10 minutes with 250 ml 1% phosphoric acid (10 ml phosphoric acid per liter ddH<sub>2</sub>O).
10. Count filter mats with beta plate reader.

15

### CELLULAR/BIOLOGIC ASSAYS

#### **EXAMPLE 22: PDGF-Induced BrdU Incorporation Assay**

20

##### Materials and Reagents:

1. PDGF: human PDGF B/B; 1276-956, Boehringer Mannheim, Germany
2. BrdU Labeling Reagent: 10 mM, in PBS (pH7.4), Cat. No. 1 647 229, Boehringer Mannheim, Germany.
- 25 3. FixDenat: fixation solution (ready to use), Cat. No. 1 647 229, Boehringer Mannheim, Germany.
4. Anti-BrdU-POD: mouse monoclonal antibody conjugated with peroxidase, Cat. No. 1 647 229, Boehringer Mannheim, Germany.
5. TMB Substrate Solution: tetramethylbenzidine (TMB), ready to use, Cat. No. 1
- 30 647 229, Boehringer Mannheim, Germany.

6. PBS Washing Solution: 1X PBS, pH 7.4, made in house (Sugen, Inc., Redwood City, California).
7. Albumin, Bovine (BSA): Fraction V powder; A-8551, Sigma Chemical Co., USA.
- 5 8. 3T3 cell line genetically engineered to express human PDGF-R.

Protocol:

1. Cells are seeded at 8000 cells/well in DMEM, 10% CS, 2 mM Gln in a 96 well plate. Cells are incubated overnight at 37 °C in 5% CO<sub>2</sub>.
- 10 2. After 24 hours, the cells are washed with PBS, and then are serum starved in serum free medium (0% CS DMEM with 0.1% BSA) for 24 hours.
3. On day 3, ligand (PDGF, 3.8 nM, prepared in DMEM with 0.1% BSA) and test compounds are added to the cells simultaneously. The negative control wells receive serum free DMEM with 0.1% BSA only; the positive control cells receive the ligand (PDGF) but no test  
15 compound. Test compounds are prepared in serum free DMEM with ligand in a 96 well plate, and serially diluted for 7 test concentrations.
4. After 20 hours of ligand activation, diluted BrdU labeling reagent (1:100 in DMEM, 0.1% BSA) is added and the cells are incubated with BrdU (final concentration=10 µM) for 1.5 hours.
- 20 5. After incubation with labeling reagent, the medium is removed by decanting and tapping the inverted plate on a paper towel. FixDenat solution is added (50 µl/well) and the plates are incubated at room temperature for 45 minutes on a plate shaker.
6. The FixDenat solution is thoroughly removed by decanting and tapping the inverted plate on a paper towel. Milk is added (5% dehydrated milk in PBS, 200 µl/well) as a  
25 blocking solution and the plate is incubated for 30 minutes at room temperature on a plate shaker.
7. The blocking solution is removed by decanting and the wells are washed once with PBS. Anti-BrdU-POD solution (1:100 dilution in PBS, 1% BSA) is added (100 µl/well) and the plate is incubated for 90 minutes at room temperature on a plate shaker.



8. The antibody conjugate is thoroughly removed by decanting and rinsing the wells 5 times with PBS, and the plate is dried by inverting and tapping on a paper towel.

9. TMB substrate solution is added (100  $\mu$ l/well) and incubated for 20 minutes at room temperature on a plate shaker until color development is sufficient for photometric  
5 detection.

10. The absorbence of the samples are measured at 410 nm (in "dual wavelength" mode with a filter reading at 490 nm, as a reference wavelength) on a Dynatech ELISA plate reader.

#### 10 **EXAMPLE 23: EGF-Induced BrdU Incorporation Assay**

##### Materials and Reagents

1. EGF: mouse EGF, 201; Toyobo, Co., Ltd. Japan
2. BrdU Labeling Reagent: 10 mM, in PBS (pH7.4), Cat. No. 1 647 229,  
15 Boehringer Mannheim, Germany.
3. FixDenat: fixation solution (ready to use), Cat. No. 1 647 229, Boehringer Mannheim, Germany.
4. Anti-BrdU-POD: mouse monoclonal antibody conjugated with peroxidase, Cat. No. 1 647 229, Boehringer Mannheim, Germany.
- 20 5. TMB Substrate Solution: tetramethylbenzidine (TMB), ready to use, Cat. No. 1 647 229, Boehringer Mannheim, Germany.
6. PBS Washing Solution : 1X PBS, pH 7.4.
7. Albumin, Bovine (BSA): Fraction V powder; A-8551, Sigma Chemical Co., USA.
- 25 8. 3T3 cell line genetically engineered to express human EGF-R.

##### Protocol

1. Cells are seeded at 8000 cells/well in 10% CS, 2mM Gln in DMEM, in a 96 well plate. Cells are incubated overnight at 37 °C in 5% CO<sub>2</sub>.
- 30 2. After 24 hours, the cells are washed with PBS, and then are serum starved in serum free medium (0% CS DMEM with 0.1% BSA) for 24 hours.

3. On day 3, ligand (EGF, 2 nM, prepared in DMEM with 0.1% BSA) and test compounds are added to the cells simultaneously. The negative control wells receive serum free DMEM with 0.1% BSA only; the positive control cells receive the ligand (EGF) but no test compound. Test compounds are prepared in serum free DMEM with ligand in a 96 well plate,  
5 and serially diluted for 7 test concentrations.
4. After 20 hours of ligand activation, diluted BrdU labeling reagent (1:100 in DMEM, 0.1% BSA) is added and the cells are incubated with BrdU (final concentration=10  $\mu$ M) for 1.5 hours.
5. After incubation with labeling reagent, the medium is removed by decanting and  
10 tapping the inverted plate on a paper towel. FixDenat solution is added (50  $\mu$ l/well) and the plates are incubated at room temperature for 45 minutes on a plate shaker.
6. The FixDenat solution is thoroughly removed by decanting and tapping the inverted plate on a paper towel. Milk is added (5% dehydrated milk in PBS, 200  $\mu$ l/well) as a blocking solution and the plate is incubated for 30 minutes at room temperature on a plate  
15 shaker.
7. The blocking solution is removed by decanting and the wells are washed once with PBS. Anti-BrdU-POD solution (1:100 dilution in PBS, 1% BSA) is added (100  $\mu$ l/well) and the plate is incubated for 90 minutes at room temperature on a plate shaker.
8. The antibody conjugate is thoroughly removed by decanting and rinsing the wells  
20 5 times with PBS, and the plate is dried by inverting and tapping on a paper towel.
9. TMB substrate solution is added (100  $\mu$ l/well) and incubated for 20 minutes at room temperature on a plate shaker until color development is sufficient for photometric detection.
10. The absorbence of the samples are measured at 410 nm (in "dual wavelength"  
25 mode with a filter reading at 490 nm, as a reference wavelength) on a Dynatech ELISA plate reader.

**EXAMPLE 24: EGF-Induced HER2-Driven BrdU Incorporation**

30 **Materials and Reagents:**

1. EGF: mouse EGF, 201; Toyobo, Co., Ltd. Japan

2. BrdU Labeling Reagent: 10 mM, in PBS (pH 7.4), Cat. No. 1 647 229, Boehringer Mannheim, Germany.
  3. FixDenat: fixation solution (ready to use), Cat. No. 1 647 229, Boehringer Mannheim, Germany.
  - 5 4. Anti-BrdU-POD: mouse monoclonal antibody conjugated with peroxidase, Cat. No. 1 647 229, Boehringer Mannheim, Germany.
  5. TMB Substrate Solution: tetramethylbenzidine (TMB), ready to use, Cat. No. 1 647 229, Boehringer Mannheim, Germany.
  6. PBS Washing Solution : 1X PBS, pH 7.4, made in house.
  - 10 7. Albumin, Bovine (BSA): Fraction V powder; A-8551, Sigma Chemical Co., USA.
  8. 3T3 cell line engineered to express a chimeric receptor having the extra-cellular domain of EGF-R and the intra-cellular domain of HER2.
- 15 Protocol:
1. Cells are seeded at 8000 cells/well in DMEM, 10% CS, 2 mM Gln in a 96- well plate. Cells are incubated overnight at 37 °C in 5% CO<sub>2</sub>.
  2. After 24 hours, the cells are washed with PBS, and then are serum starved in serum free medium (0% CS DMEM with 0.1% BSA) for 24 hours.
  - 20 3. On day 3, ligand (EGF=2 nM, prepared in DMEM with 0.1% BSA) and test compounds are added to the cells simultaneously. The negative control wells receive serum free DMEM with 0.1% BSA only; the positive control cells receive the ligand (EGF) but no test compound. Test compounds are prepared in serum free DMEM with ligand in a 96 well plate, and serially diluted for 7 test concentrations.
  - 25 4. After 20 hours of ligand activation, diluted BrdU labeling reagent (1:100 in DMEM, 0.1% BSA) is added and the cells are incubated with BrdU (final concentration = 10 µM) for 1.5 hours.
  5. After incubation with labeling reagent, the medium is removed by decanting and tapping the inverted plate on a paper towel. FixDenat solution is added (50 µl/well) and the  
30 plates are incubated at room temperature for 45 minutes on a plate shaker.

6. The FixDenat solution is thoroughly removed by decanting and tapping the inverted plate on a paper towel. Milk is added (5% dehydrated milk in PBS, 200  $\mu$ l/well) as a blocking solution and the plate is incubated for 30 minutes at room temperature on a plate shaker.
- 5 7. The blocking solution is removed by decanting and the wells are washed once with PBS. Anti-BrdU-POD solution (1:100 dilution in PBS, 1% BSA) is added (100  $\mu$ l/well) and the plate is incubated for 90 minutes at room temperature on a plate shaker.
8. The antibody conjugate is thoroughly removed by decanting and rinsing the wells 5 times with PBS, and the plate is dried by inverting and tapping on a paper towel.
- 10 9. TMB substrate solution is added (100  $\mu$ l/well) and incubated for 20 minutes at room temperature on a plate shaker until color development is sufficient for photometric detection.
- 15 10. The absorbance of the samples are measured at 410 nm (in "dual wavelength" mode with a filter reading at 490 nm, as a reference wavelength) on a Dynatech ELISA plate reader.

#### **EXAMPLE 25: IGF1-Induced BrdU Incorporation Assay**

##### **Materials and Reagents:**

- 20 1. IGF1 Ligand: human, recombinant; G511, Promega Corp, USA.
2. BrdU Labeling Reagent: 10 mM, in PBS (pH 7.4), Cat. No. 1 647 229, Boehringer Mannheim, Germany.
3. FixDenat: fixation solution (ready to use), Cat. No. 1 647 229, Boehringer Mannheim, Germany.
- 25 4. Anti-BrdU-POD: mouse monoclonal antibody conjugated with peroxidase, Cat. No. 1 647 229, Boehringer Mannheim, Germany.
5. TMB Substrate Solution: tetramethylbenzidine (TMB), ready to use, Cat. No. 1 647 229, Boehringer Mannheim, Germany.
6. PBS Washing Solution: 1X PBS, pH 7.4.
- 30 7. Albumin, Bovine (BSA): Fraction V powder; A-8551, Sigma Chemical Co., USA.

8. 3T3 cell line genetically engineered to express human IGF-1 receptor.

Protocol:

1. Cells are seeded at 8000 cells/well in DMEM, 10% CS, 2 mM Gln in a 96- well  
5 plate. Cells are incubated overnight at 37 °C in 5% CO<sub>2</sub>.
2. After 24 hours, the cells are washed with PBS, and then are serum starved in  
serum free medium (0% CS DMEM with 0.1% BSA) for 24 hours.
3. On day 3, ligand (IGF1=3.3 nM, prepared in DMEM with 0.1% BSA) and test  
10 compounds are added to the cells simultaneously. The negative control wells receive serum free  
DMEM with 0.1% BSA only; the positive control cells receive the ligand (IGF1) but no test  
compound. Test compounds are prepared in serum free DMEM with ligand in a 96 well plate,  
and serially diluted for 7 test concentrations.
4. After 16 hours of ligand activation, diluted BrdU labeling reagent (1:100 in  
15 DMEM, 0.1% BSA) is added and the cells are incubated with BrdU (final concentration=10 µM)  
for 1.5 hours.
5. After incubation with labeling reagent, the medium is removed by decanting and  
tapping the inverted plate on a paper towel. FixDenat solution is added (50 µl/well) and the  
plates are incubated at room temperature for 45 minutes on a plate shaker.
6. The FixDenat solution is thoroughly removed by decanting and tapping the  
20 inverted plate on a paper towel. Milk is added (5% dehydrated milk in PBS, 200 µl/well) as a  
blocking solution and the plate is incubated for 30 minutes at room temperature on a plate  
shaker.
7. The blocking solution is removed by decanting and the wells are washed once  
with PBS. Anti-BrdU-POD solution (1:100 dilution in PBS, 1% BSA) is added (100 µl/well)  
25 and the plate is incubated for 90 minutes at room temperature on a plate shaker.
8. The antibody conjugate is thoroughly removed by decanting and rinsing the wells  
5 times with PBS, and the plate is dried by inverting and tapping on a paper towel.
9. TMB substrate solution is added (100 µl/well) and incubated for 20 minutes at  
room temperature on a plate shaker until color development is sufficient for photometric  
30 detection.

10. The absorbence of the samples are measured at 410 nm (in "dual wavelength" mode with a filter reading at 490 nm, as a reference wavelength) on a Dynatech ELISA plate reader.

5 **EXAMPLE 26: HUV-EC-C Assay**

The following protocol may also be used to measure a compound's activity against PDGF-R, FGF-R, VEGF, aFGF or Flk-1/KDR, all of which are naturally expressed by HUV-EC cells.

10

**DAY 0**

1. Wash and trypsinize HUV-EC-C cells (human umbilical vein endothelial cells, (American Type Culture Collection; catalogue no. 1730 CRL). Wash with Dulbecco's phosphate-buffered saline (D-PBS; obtained from Gibco BRL; catalogue no. 14190-029) 2 times  
15 at about 1 ml/10 cm<sup>2</sup> of tissue culture flask. Trypsinize with 0.05% trypsin-EDTA in non-enzymatic cell dissociation solution (Sigma Chemical Company; catalogue no. C-1544). The 0.05% trypsin was made by diluting 0.25% trypsin/1 mM EDTA (Gibco; catalogue no. 25200-049) in the cell dissociation solution. Trypsinize with about 1 ml/25-30 cm<sup>2</sup> of tissue culture flask for about 5 minutes at 37 °C. After cells have detached from the flask, add an equal  
20 volume of assay medium and transfer to a 50 ml sterile centrifuge tube (Fisher Scientific; catalogue no. 05-539-6).

2. Wash the cells with about 35 ml assay medium in the 50 ml sterile centrifuge tube by adding the assay medium, centrifuge for 10 minutes at approximately 200 g, aspirate the supernatant, and resuspend with 35 ml D-PBS. Repeat the wash two more times with D-PBS,  
25 resuspend the cells in about 1 ml assay medium/15 cm<sup>2</sup> of tissue culture flask. Assay medium consists of F12K medium (Gibco BRL; catalogue no. 21127-014) + 0.5% heat-inactivated fetal bovine serum. Count the cells with a Coulter Counter™ (Coulter Electronics, Inc.) and add assay medium to the cells to obtain a concentration of 0.8-1.0x10<sup>5</sup> cells/ml.

3. Add cells to 96-well flat-bottom plates at 100 µl/well or 0.8-1.0x10<sup>4</sup> cells/well;  
30 incubate ~24 h at 37 °C, 5% CO<sub>2</sub>.

DAY 1

1. Make up two-fold drug titrations in separate 96-well plates, generally 50  $\mu$ M on down to 0  $\mu$ M. Use the same assay medium as mentioned in day 0, step 2, above. Titrations are made by adding 90  $\mu$ l/well of drug at 200  $\mu$ M (4X the final well concentration) to the top well of a particular plate column. Since the stock drug concentration is usually 20 mM in DMSO, the 200  $\mu$ M drug concentration contains 2% DMSO.

- Therefore, diluent made up to 2% DMSO in assay medium (F12K + 0.5% fetal bovine serum) is used as diluent for the drug titrations in order to dilute the drug but keep the DMSO concentration constant. Add this diluent to the remaining wells in the column at 60  $\mu$ l/well.
- 10 Take 60  $\mu$ l from the 120  $\mu$ l of 200  $\mu$ M drug dilution in the top well of the column and mix with the 60  $\mu$ l in the second well of the column. Take 60  $\mu$ l from this well and mix with the 60  $\mu$ l in the third well of the column, and so on until two-fold titrations are completed. When the next-to-the-last well is mixed, take 60  $\mu$ l of the 120  $\mu$ l in this well and discard it. Leave the last well with 60  $\mu$ l of DMSO/media diluent as a non-drug-containing control. Make 9 columns of
- 15 titrated drug, enough for triplicate wells each for 1) VEGF (obtained from Pepro Tech Inc., catalogue no. 100-200, 2) endothelial cell growth factor (ECGF) (also known as acidic fibroblast growth factor, or aFGF) (obtained from Boehringer Mannheim Biochemica, catalogue no. 1439 600); or, 3) human PDGF B/B (1276-956, Boehringer Mannheim, Germany) and assay media control. ECGF comes as a preparation with sodium heparin.

- 20 2. Transfer 50  $\mu$ l/well of the drug dilutions to the 96-well assay plates containing the  $0.8-1.0 \times 10^4$  cells/100  $\mu$ l/well of the HUV-EC-C cells from day 0 and incubate  $\sim 2$  h at 37  $^{\circ}$ C, 5% CO<sub>2</sub>.

3. In triplicate, add 50  $\mu$ l/well of 80  $\mu$ g/ml VEGF, 20 ng/ml ECGF, or media control to each drug condition. As with the drugs, the growth factor concentrations are 4X the desired
- 25 final concentration. Use the assay media from day 0, step 2, to make the concentrations of growth factors. Incubate approximately 24 hours at 37  $^{\circ}$ C, 5% CO<sub>2</sub>. Each well will have 50  $\mu$ l drug dilution, 50  $\mu$ l growth factor or media, and 100  $\mu$ l cells, = 200  $\mu$ l /well total. Thus the 4X concentrations of drugs and growth factors become 1X once everything has been added to the wells.

DAY 2

1. Add  $^3\text{H}$ -thymidine (Amersham; catalogue no. TRK-686) at 1  $\mu\text{Ci}/\text{well}$  (10  $\mu\text{l}/\text{well}$  of 100  $\mu\text{Ci}/\text{ml}$  solution made up in RPMI media + 10% heat-inactivated fetal bovine serum) and incubate ~24 h at 37 °C, 5%  $\text{CO}_2$ . Note:  $^3\text{H}$ -thymidine is made up in RPMI media because all of the other applications for which we use the  $^3\text{H}$ -thymidine involve experiments done in RPMI. The media difference at this step is probably not significant. RPMI was obtained from Gibco BRL, catalogue no. 11875-051.

10 DAY 3

1. Freeze plates overnight at -20°C.

DAY 4

1. Thaw plates and harvest with a 96-well plate harvester (Tomtec Harvester 96<sup>(R)</sup>) onto filter mats (Wallac; catalogue no. 1205-401); read counts on a Wallac Betaplate<sup>(TM)</sup> liquid scintillation counter.



### CONCLUSION

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The molecular complexes and the methods, procedures, treatments, molecules, specific compounds described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

10 All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

The invention illustratively described herein suitably may be practiced in the absence of  
15 any element or elements, limitation or limitations that are not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any  
20 equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are  
25 considered to be within the scope of this invention as defined by the appended claims.

In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine, chlorine, and iodine, claims for X  
30 being bromine and claims for X being bromine and chlorine are fully described.

In view of the degeneracy of the genetic code, other combinations of nucleic acids also encode the claimed peptides and proteins of the invention. For example, all four nucleic acid sequences GCT, GCC, GCA, and GCG encode the amino acid alanine. Therefore, if for an amino acid there exists an average of three codons, a polypeptide of 100 amino acids in length will, on average, be encoded by 3100, or  $5 \times 10^4$ , nucleic acid sequences. Thus, a nucleic acid sequence can be modified to form a second nucleic acid sequence, encoding the same polypeptide as encoded by the first nucleic acid sequences, using routine procedures and without undue experimentation. Thus, all possible nucleic acids that encode the claimed peptides and proteins are also fully described herein, as if all were written out in full taking into account the codon usage, especially that preferred in humans. Furthermore, changes in the amino acid sequences of polypeptides, or in the corresponding nucleic acid sequence encoding such polypeptide, may be designed or selected to take place in an area of the sequence where the significant activity of the polypeptide remains unchanged. For example, an amino acid change may take place within a  $\beta$ -turn, away from the active site of the polypeptide. Also changes such as deletions (*e.g.* removal of a segment of the polypeptide, or in the corresponding nucleic acid sequence encoding such polypeptide, which does not affect the active site) and additions (*e.g.* addition of more amino acids to the polypeptide sequence without affecting the function of the active site, such as the formation of GST-fusion proteins, or additions in the corresponding nucleic acid sequence encoding such polypeptide without affecting the function of the active site) are also within the scope of the present invention. Such changes to the polypeptides can be performed by those with ordinary skill in the art using routine procedures and without undue experimentation. Thus, all possible nucleic and/or amino acid sequences that can readily be determined not to affect a significant activity of the peptide or protein of the invention are also fully described herein.

The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

Other embodiments are within the following claims.

What is claimed is:

CLAIMS

1. An isolated, enriched or purified nucleic acid molecule encoding a kinase  
5 polypeptide, wherein said nucleic acid molecule comprises a nucleotide sequence that:  
(a) encodes a polypeptide having an amino acid sequence selected from the group consisting  
of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID  
NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67,  
SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID  
10 NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ  
ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84,  
SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID  
NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ  
ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101,  
15 SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ  
ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID  
NO:112, SEQ ID NO:113, and SEQ ID NO:114;  
(b) is the complement of the nucleotide sequence of (a);  
(c) hybridizes under stringent conditions to the nucleotide molecule of (a) and  
20 encodes a naturally occurring kinase polypeptide;  
(d) encodes a polypeptide having an amino acid sequence selected from the group consisting  
of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID  
NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67,  
SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID  
25 NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ  
ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84,  
SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID  
NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ  
ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101,  
30 SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ  
ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID

NO:112, SEQ ID NO:113, and SEQ ID NO:114, except that it lacks one or more, but not all, of an N-terminal domain, a C-terminal catalytic domain, a catalytic domain, a C-terminal domain, a coiled-coil structure region, a proline-rich region, a spacer region and a C-terminal tail; or

(e) is the complement of the nucleotide sequence of (d).

5

2. The nucleic acid molecule of claim 1, further comprising a vector or promoter effective to initiate transcription in a host cell.

10

3. The nucleic acid molecule of claim 1, wherein said nucleic acid molecule is isolated, enriched, or purified from a mammal.

15

4. The nucleic acid molecule of claim 3, wherein said mammal is a human.

5. The nucleic acid probe of claim 1 used for the detection of nucleic acid encoding a kinase polypeptide in a sample, wherein said kinase polypeptide is selected from the group consisting of a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114.

25

30

6. A recombinant cell comprising the nucleic acid molecule of claim 1 encoding a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID  
 5 NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96,  
 10 SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114.

15

7. An isolated, enriched, or purified kinase polypeptide, wherein said polypeptide comprises an amino acid sequence having
- (a) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ  
 20 ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID  
 25 NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, respectively;
- 30 (b) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ

ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, and SEQ ID NO: 114, respectively, except that it lacks one or more, but not all, of the domains selected from the group consisting of an N-terminal domain, a C-terminal catalytic domain, a catalytic domain, a C-terminal domain, a coiled-coil structure region, a proline-rich region, a spacer region, and a C-terminal tail.

15

8. The kinase polypeptide of claim 7, wherein said polypeptide is isolated, purified, or enriched from a mammal.

20

9. The kinase polypeptide of claim 8, wherein said mammal is a human.

10. An antibody or antibody fragment having specific binding affinity to a kinase polypeptide or to a domain of said polypeptide, wherein said polypeptide is a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ

ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114.

5

11. A hybridoma which produces an antibody having specific binding affinity to a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114.

20

12. A kit comprising an antibody which binds to a polypeptide of claim 7 or 8 and negative control antibody

25

13. A method for identifying a substance that modulates the activity of a kinase polypeptide comprising the steps of:

(a) contacting the kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ

ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, 5 SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114 with a test substance;

(b) measuring the activity of said polypeptide; and

(c) determining whether said substance modulates the activity of said polypeptide.

10

14. A method for identifying a substance that modulates the activity of a kinase polypeptide in a cell comprising the steps of:

(a) expressing a kinase polypeptide having an amino acid sequence selected from the 15 group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, 20 SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114; 25

(b) adding a test substance to said cell; and

(c) monitoring a change in cell phenotype or the interaction between said polypeptide and a natural binding partner.

30



15. A method for treating a disease or disorder by administering to a patient in need of such treatment a substance that modulates the activity of a kinase having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114.

15

16. The method of claim 15, wherein said disease or disorder is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders.

20

17. The method of claim 15, wherein said disease or disorder is selected from the group consisting of cancers of tissues; cancers of hematopoietic origin; diseases of the central nervous system; diseases of the peripheral nervous system; Alzheimer's disease; Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis; viral infections; infections caused by prions; infections caused by bacteria; infections caused by fungi; and ocular diseases.

25

18. The method of claim 15, wherein said disease or disorder is selected from the group consisting of migraines; pain; sexual dysfunction; mood disorders; attention disorders; cognition disorders; hypotension; hypertension; psychotic disorders; neurological disorders; dyskinesias; metabolic disorders; and organ transplant rejection.

30

19. The method of claim 15, wherein said substance modulates kinase activity *in vitro*.

5

20. The method of claim 19, wherein said substance is a kinase inhibitor.

21. A method for detection of a kinase polypeptide in a sample as a diagnostic tool  
10 for a disease or disorder, wherein said method comprises:  
(a) contacting said sample with a nucleic acid probe which hybridizes under hybridization  
assay conditions to a nucleic acid target region of a kinase polypeptide having an amino acid  
sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID  
NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ  
15 ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70,  
SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID  
NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ  
ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87,  
SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID  
20 NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ  
ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID  
NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID  
NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID  
NO:114, said probe comprising the nucleic acid sequence encoding said polypeptide, fragments  
25 thereof, or the complements of said sequences and fragments; and  
(b) detecting the presence or amount of the probe:target region hybrid as an indication of  
said disease.

30 22. The method of claim 21, wherein said disease or disorder is selected from the  
group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain  
or neuronal-associated diseases, and metabolic disorders.

23. The method of claim 21, wherein said disease or disorder is selected from the group consisting of cancers of tissues; cancers of hematopoietic origin; diseases of the central nervous system; diseases of the peripheral nervous system; Alzheimer's disease; Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis; viral infections; infections caused by prions; infections caused by bacteria; infections caused by fungi; and ocular diseases.

24. The method of claim 21, wherein said disease or disorder is selected from the group consisting of migraines, pain; sexual dysfunction; mood disorders; attention disorders; cognition disorders; hypotension; hypertension; psychotic disorders; neurological disorders; dyskinesias; metabolic disorders; and organ transplant rejection.

25. A method for detection of a kinase polypeptide in a sample as a diagnostic tool for a disease or disorder, wherein said method comprises:

(a) comparing a nucleic acid target region encoding said kinase polypeptide in a sample, wherein said kinase polypeptide has an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, and SEQ ID NO:114, or one or more fragments thereof, with a control nucleic acid target region encoding said kinase polypeptide, or one or more fragments thereof; and

(b) detecting differences in sequence or amount between said target region and said control target region, as an indication of said disease or disorder.

5           26.     The method of claim 25, wherein said disease or disorder is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders.

10           27.     The method of claim 25, wherein said disease or disorder is selected from the group consisting of cancers of tissues; cancers of hematopoietic origin; diseases of the central nervous system; diseases of the peripheral nervous system; Alzheimer's disease; Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis; viral infections; infections caused by prions; infections caused by bacteria; infections caused by fungi; and ocular diseases.

15           28.     The method of claim 25, wherein said disease or disorder is selected from the group consisting of migraines, pain; sexual dysfunction; mood disorders; attention disorders; cognition disorders; hypotension; hypertension; psychotic disorders; neurological disorders;  
20     dyskinesias; metabolic disorders; and organ transplant rejection.

&gt;SGK187\_CRIK\_ID#NA\_1

ATGTTGAAGTTCAAATATGGAGCGCGGAATCCTTTGGATGCTGGTGTCTGCTGAACCCATTGCCAGCCGGGC  
CTCCAGGCTGAATCTGTTCTTCCAGGGGAAACCACCTTTATGACTCAACAGCAGATGTCTCCTCTTTCCC  
GAGAAGGGATATTAGATGCCCTCTTTGTTCTCTTTGAAGAATGCAGTCAGCCTGCTCTGATGAAGATTAAG  
CACGTGAGCAACTTTGTCCCAGGAAGTGATTTCCGACACCATAGCTGAGTTACAGGAGCTCCAGCCTTCGGC  
AAAGGACTTCGAAGTCAGAAGTCTTGTAGGTTGTGGTCACTTTGCTGAAGTGCAGGTGGTAAGAGAGAAAG  
CAACCGGGGACATCTATGCTATGAAAGTGATGAAGAAGAAGGCTTTATTGGCCCAGGAGCAGGTTTCATTT  
TTTGAGGAAGAGCGGAACATATTATCTCGAAGCACAAGCCCGTGGATCCCCCAATTACAGTATGCCTTTCA  
GGACAAAAATCACCTTTATCTGGTCATGGAATATCAGCCTGGAGGGGACTTGCTGTCACTTTTGAATAGAT  
ATGAGGACCAGTTAGATGAAAACCTGATACAGTTTACCTAGCTGAGCTGATTTTGGCTGTTACAGCGTT  
CATCTGATGGGATACGTGCATCGAGACATCAAGCCTGAGAACATTCTCGTTGACCGCAGGACACATCAAA  
GCTGGTGGATTTTGGATCTGCCCGGAAAATGAAATCAAACAAGATGGTGAATGCCAAACTCCCGATTGGGA  
CCCCAGATTACATGGCTCCTGAAGTGTGACTGTGATGAACGGGGATGGAAAAGCACCTACGGCCTGGAC  
TGTGACTGGTGGTCACTGGGCGTGATTGCCATGAGATGATTTATGGGAGATCCCCCTTCGCAGAGGGAAC  
CTCTGCCAGAACCTTCAATAACATTATGAATTTCCAGCGGTTTTTGAAATTTCCAGATGACCCCAAAGTGA  
GCAGTGACTTTCTTGATCTGATTCAAAGCTTGTGTGCGGCCAGAAAGAGAGACTGAAGTTTGAAGGTCTT  
TGCTGCCATCCTTTCTCTCTAAAATTGACTGGAACAACATTCTGTAACCTCCTCCCCCTTCGTTCCAC  
CCTCAAGTCTGACGATGACACCTCCAATTTTGATGAACCAGAGAAGAATTCGTGGGTTTCATCTCTCCGT  
GCCAGCTGAGCCCCCAGGCTTCTCGGTGAAGAACTGCCGTTTGTGGGGTTTTCGTACAGCAAGGCACCTG  
GGGATTCTTGGTAGATCTGAGTCTGTTGTGTGCGGCTCGGACTCCCTGCCAAGACTAGCTCCATGGAAAA  
GAACTTCTCATCAAAAGCAAAGAGCTACAAGACTCTCAGGACAAGTGTCAAGATGGAGCAGGAAATGA  
CCCGGTTACATCGGAGAGTGTGAGAGGTGGAGGCTGTGCTTAGTCAGAAGGAGGTGGAGCTGAAGGCCTCT  
GAGACTCAGAGATCCCTCCTGGAGCAGGACCTTGCTACCTACATCACAGAATGCAGTAGCTTAAAGCGAAG  
TTTGGAGCAAGCACGGATGGAGGTGTCCAGGAGGATGACAAAGCACTGCAGCTTCTCCATGATATCAGAG  
AGCAGAGCCGGAAGCTCCAAGAAATCAAAGAGCAGGAGTACCAGGCTCAAGTGGAGAAATGAGGTTGATG  
ATGAATCAGTTGGAAGAGGATCTTGTCTCAGCAAGAAGACGGAGTGATCTCTACGAATCTGAGCTGAGAGA  
GTCTCGGCTTGCTGCTGAAGAATTCAAGCGGAAAGCGACAGAATGTGAGCATAACTGTTGAAGGCTAAGG  
ATCAAGGGAAGCCTGAAGTGGGAGAATATGCGAACTGGAGAAGATCAATGCTGAGCAGCAGCTCAAAATT  
CAGGAGCTCCAAGAGAAACTGGAGAGGCTGTAAAGCGACGACGAGGCCACCGAGGATTCTTCTGAAGGCA  
CCGCCAGGCAAAGGAGCGAGCCGAGAGGGAGCTGGAGAAGCTGCAGAAACCGAGAGGATTCTTCTGAAGGCA  
TCAGAAAGAAGCTGGTGGAAAGCTGAGGAACGCCGCCATTCTCTGGAGAACAAGGTAAAGAGACTAGAGACC  
ATGGAGCGTAGAGAAAAACAGACTGAAGGATGACATCCAGACAAAATCCCAACAGATCCAGCAGATGGCTGA  
TAAATTTCTGGAGCTCGAAGAGAAACATCGGGAGGCCCAAGTCTCAGCCCAGCACCTAGAAGTGACCTGA  
AACAGAAAGAGCAGCACTATGAGGAAAAGATTAAAGTGTTGGACAATCAGATAAAGAAAGACCTGGCTGAC  
AAGGAGACACTGGAGAACATGATGCAGAGACACGAGGAGGAGGCCATGAGAAGGGCAAATTTCTCAGCGA  
ACAGAAGGCGATGATCAATGCTATGGATTCCAAGATCAGATCCCTGGAACAGAGGATTGTGGAAGTGTCTG  
AAGCCAAATAACTTGCAGCAAATAGCAGTCTTTTACCCAAAGGAACATGAAGGCCCAAGAAGAGATGATT  
TCTGAACTCAGGCAACAGAAATTTTACCTGGAGACACAGGCTGGGAAGTTGGAGGCCCAGAACCGAAAAC  
GGAGGAGCAGCTGGAGAAGATCAGCCACCAAGACCACAGTGACAAGAATCGGCTGCTGGAAGTGGAGACAA  
GATTGCGGGAGGTGAGTCTAGAGCACGAGGAGCAGAACTGGAGCTCAAGCGCCAGCTCACAGAGCTACAG  
CTCTCCCTGCAGGAGCGCGAGTCACAGTTGACAGCCCTGCAGGCTGCACGGGCGGCCCTGGAGAGCCAGCT  
TCGCCAGGCGAAGACAGAGCTGGAAGAGACCACAGCAGAAGCTGAAGAGGAGATCCAGGCACTCACGGCAC  
ATAGAGATGAAATCCAGCGCAAATTTGATGCTCTTCGTAACAGCTGTACTGTAATCACAGACCTGGAGGAG  
CAGCTAAACCAGCTGACCGAGGACAACGCTGAACTCAACAACCAAACTTCTACTTGTCCAAACAACCTCGA  
TGAGGCTTCTGGCGCCAACGACGAGATTGTACAACTGCGAAGTGAAGTGGACCATCTCCGCGGGAGATCA  
CGGAACGAGAGATGCAGCTTACCAGCCAGAAGCAACGATGGAGGCTCTGAAGACCACGTCACCATGCTG  
GAGGAACAGGTGATGGATTTGGAGGCCCTAAACGATGAGCTGCTAGAAAAAGAGCGCGCAGTGGGAGGCCCTG  
GAGGAGCGTCTGGGTGATGAGAAATCCAGTTTGAGTGTGCGGTTTCGAGAGCTGCAGAGAATGCTGGACA  
CCGAGAAACAGAGCAGGGCGAGAGCCGATCAGCGGATCACCGAGTCTCGCCAGGTGGTGGAGCTGGCAGTG  
AAGGAGCACAAGGCTGAGATTCTCGCTCTGCAGCAGGCTCTCAAAGAGCAGAAGCTGAAGGCCGAGAGCCT  
CTCTGACAAGCTCAATGACCTGGAGAAGAAGCATGCTATGCTTGAATGAATGCCCGAAGCTTACAGCAGA  
AGCTGGAGACTGAACGAGAGCTCAAACAGAGGCTTCTGGAAGAGCAAGCCAAATTACAGCAGCAGATGGAC  
CTGCAGAAAAATCACATTTTCCGTCTGACTCAAGGACTGCAAGAAGCTCTAGATCGGGCTGATCTACTGAA

Figure 1A

GACAGAAAGAAGTGACTTGGAGTATCAGCTGGAAAAACATTTCAGGTTCTCTATTCTCATGAAAAGGTGAAAA  
 TGAAGGCACTATTTCTCAACAAACCAAACTCATTGATTTTCTGCAAGCCAAAATGGACCAACCTGCTAAA  
 AAGAAAAAGGTTCTCTGCAGTACAAATGAGCTGAAGCTGGCCCTGGAGAAGGAGAAAGCTCGCTGTGCAGA  
 GCTAGAGGAAGCCCTTCAGAAGACCCGCATCGAGCTCCGGTCCGCCCCGGGAGGAAGCTGCCACCGCAAAG  
 CAACGGACCACCCACACCCATCCACGCCAGCCACCGCGAGGCAGCAGATCGCCATGTCCGCCATCGTGCCG  
 TCGCCAGAGCACCAGCCAGTGCATGAGCCTGTGGCCCCGCCATCCAGCCGCAGAAAGGAGTCTTCAAC  
 TCCAGAGGAATTTAGTCGGCGTCTTAAGGAACGCATGCACCACAATATTCTCACCAGTTCAACGTAGGAC  
 TGAACATGCGAGGCCACAAAGTGTCTGTGTGTCTGGATACCGTGCACCTTTGGACGCCAGGCATCCAAATGT  
 CTCGAATGTTCAGGTGATGTGTCAACCCCAAGTGCTCCACGTGCTTGCCAGCCACCTGCGGCTTGCTGTGA  
 ATATGCCACACACTTCACCGAGGCCTTCTGCCGTGACAAAATGAACTCCCCAGGTCTCCAGACCAAGGAGC  
 CCAGCAGCAGCTTGCACCTGGAAGGGTGGATGAAGGTGCCCAGGAATAACAAACGAGGACAGCAAGGCTGG  
 GACAGGAAGTACATTGTCTGGAGGGATCAAAAGTCTCATTATGACAATGAAGCCAGAGAAGCTGGACA  
 GAGGCCGGTGAAGAATTTGAGCTGTGCCTTCCCGACGGGGATGTATCTATTTCATGGTGCCGTTGGTGCTT  
 CCGAACTCGCAAATACAGCCAAAGCAGATGTCCATACATACTGAAGATGGAATCTCACCCGCACACCACC  
 TGCTGGCCCCGGGAGAACCTCTACTTGTCTAGCTCCAGCTTCCCTGACAAACAGCGCTGGGTACCCGCTT  
 AGAATCAGTTGTGCGAGGTGGGAGAGTTTCTAGGAAAAAGCAGAAGCTGATGTCTAACTGCTTGGAACT  
 CCTGTCTGAACTGGAAGGTGATGACCGTCTAGACATGAACTGCACGCTGCCCTTCAGTGACCAGGTGGTG  
 TTGGTGGGCACCGAGGAAGGGCTCTACGCCCTGAATGTCTTGAAAACTCCCTAACCCATGTCCCAGGAAT  
 TGGAGCAGTCTTCCAAATTTATATTATCAAGGACCTGGAGAAGCTACTCATGATAGCAGGAGAAGAGCGGG  
 CACTGTGTCTTGTGGACGTGAAGAAAGTGAACAGTCCCTGGCCAGTCCCACCTGCCTGCCAGCCCGAC  
 ATCTCACCCAAACATTTTGAAGCTGTCAAGGGCTGCCACTTGTGTTGGGGCAGGCAAGATTGAGAACGGGCT  
 CTGCATCTGTGCAGCCATGCCAGCAAAGTCGTCTTCTCCGCTACAACGAAACCTCAGCAAATACTGCA  
 TCCGGAAAGAGATAGAGACCTCAGAGCCCTGCAGCTGTATCCACTTCACCAATTACAGTATCCTCATTGGA  
 ACCAATAAATTTCTACGAAATCGACATGAAGCAGTACACGCTCGAGGAATTCTCGGATAAGAATGACCATT  
 CTTGGCACCTGTGTGTTTGGCCCTCTTCCACAGCTTCCCTGTCTCAATCGTGCAGGTGAACAGCGCAG  
 GGCAGCGAGAGGAGTACTTGCTGTGTTTCCACGAATTTGGAGTGTTCGTGGATTCTTACGGAAGACGTAGC  
 CGCACAGACGATCTCAAGTGGAGTCGCTTACCTTTGGCCCTTTGGCTACAGAGAACCCTATCTGTTTGTGAC  
 CCACTTCAACTCACTCGAAGTAATTGAGATCCAGGCACGCTCCTCAGCAGGGACCCCTGCCCGAGCGTACC  
 TGGACATCCCGAACCCGCGCTACCTGGGCCCTGCCATTTCTCAGGAGCGATTTACTTGGCGTCTCATAC  
 CAGGATAAATTAAGGGTCATTTGTGCAAGGGAAACCTCGTGAAGGAGTCCGGCACTGAACACCACCGGGG  
 CCCGTCCACCTCCCGCAGCAGCCCCAACAAAGCGAGGCCACCCACGTACAACGAGCACATCACCAGCGCG  
 TGGCCTCCAGCCCCAGCGCCCGCCGAAGGCCCCAGCCACCCGCGAGAGCCAAGCACACCCACCGCTACCGC  
 GAGGGGCGGACCGAGCTGCGCAGGGACAAGTCTCTGCGCCGCCCTGGAGCGAGAGAAGTCCCCGGCCG  
 GATGCTCAGCACGCGGAGAGAGCGGTCCCCCGGAGGCTGTTTGAAGACAGCAGCAGGGGCCGGCTGCCTG  
 CGGGAGCCGTGAGGACCCGCTGTCCAGGTGAACAAGGTGAGGCAGCATTCC

>SGK064\_GRK7\_ID#NA\_2

ATGGTGGACATGGGGGCCCTGGACAACCTGATCGCCAACACCGCCTACCTGCAGGCCCGGAAGCCCTCGGA  
 CTGCGACAGCAAAGAGCTGCAGCGGCGGCGGTAGCCTGGCCCTGCCCGGGCTGCAGGGCTGCGCGGAGC  
 TCCGCCAGAAGCTGTCCCTGAACTTCCACAGCCTGTGTGAGCAGCAGCCCATCGGTGCGCGCCTCTTCCGT  
 GACTTCCTAGCCACAGTGGCCACGTCCGCAAGGCGGCAACCTTCCTAGAGGACGTGCAGAACTGGGAGCT  
 GGCCGAGGAGGGACCCACCAAGACAGCGCGCTGCAGGGGCTGGTGGCCACTTGTGCGAGTGGCCCTGCCC  
 CGGGGAACCCGCAACCTTCTCAGCCAGGCCGTGGCCACCAAGTGCCAAGCAGCCACCACTGAGGAAGAG  
 CGAGTGGCTGCAGTGACGCTGGCCAAGGCTGAGGCCATGGCTTTCTTGCAAGAGCAGCCCTTTAAGGATTT  
 CGTGACCAGCGCCTTCTACGACAAGTTTCTGCAGTGGAACTCTTCGAGATGCAACCAGTGTGAGACAAGT  
 ACTTCACTGAGTTTCAGAGTGTCTGGGGAAGGTGGTTTTTGGGGAGGTATGTGCCGTCCAGGTGAAAAACACT  
 GGGAAAGATGTATGCCCTGTAAGAACTGGACAAGAAGCGGCTGAAGAAGAAAGGTGGCGAGAAGATGGCTCT  
 CTTGGAAAAAGGAAATCTTGGAGAAGGTGAGAGCCCTTTCATTGTCTCTTGGCCTATGCCCTTTGAGAGCA  
 AGACCCATCTCTGCCTTGTCTATGAGCCTGATGAATGGGGGAGACCTCAAGTTCACATCTACAACGTGGGC  
 ACGCGTGGCCTGGACATGAGCCGGGTGATCTTTTACTCGGCCAGATAGCCTGTGGGATGCTGCACCTCCA  
 TGAACCTGGCATCGTCTATCGGGACATGAAGCCTGAGAATGGGCTTCTGGATGACCTCGGCAACTGCAGGT  
 TATCTGACCTGGGGCTGGCCGTGGAGATGAAGGGTGGCAAGCCCATCACCAGAGGGCTGGAACCAATGGT  
 TACATGGCTCCTGAGATCCTAATGGAAGGTAAAGTTATTCCTATCCTGTGGACTGGTTTGCCATGGGATG

Figure 1B

CAGCATTATGAAATGGTTGCTGGACGAACACCATTCAAAGATTACAAGGAAAAGGTCAGTAAAGAGGATC  
TGAAGCAAAGAACTCTGCAAGACGAGGTCAAATTCCAGCATGATAACTTCACAGAGGAAGCAAAGATATT  
TGCAGGCTCTTCTTGGCTAAGAAACCAGAGCAACGCTTAAGAAGCAGAGAAAAGTCTGATGATCCCAGGAA  
ACATCATTTCTTTAAAACGATCAACTTTCCTCGCCTGGAAGCTGGCCTAATTGAACCCCCATTTGTGCCAG  
ACCCTTCAGTGGTTTATGCCAAAGACATCGCTGAAATTGATGATTTCTCTGAGGTTCTGGGGGGTGAATTT  
GATGACAAAGATAAGCAGTTCTTCAAAAACCTTTGCGACAGGTGCTGTTCTTATAGCATGGCAGGAAGAAAT  
TATAGAAACGGGACTGTTTGAGGAACGAATGACCCCAACAGACCTACGGGTTGTGAGGAGGGTAATTCAT  
CCAAGTCTGGCGTGTGTTGTTATTGTAA

>SGK409\_KIAA0303\_ID#NA\_3

ATGCAGACAAAACACTTCCTGGTGGTGACTCATGTTACTCCAAAGCATCTCCACCCAAGCACCTGTTTCAGC  
TGTGACATCTTACACGCAGTGTTCTGTCCATGCGTGCTCACCTCCTGCCCCCTCAGGGGCAGCCAGTCCTC  
CCTCAAGTTGCCGAACAAGCAACCGGAAAAGCTTAATAGGCAATGGGCAGTCACCAGCATTGCCTCGACCA  
CACTCACCTCTCTCTGCTCATGCAGAGTCTCTTTTGGAGCCATACATGAGAATTGTTTGAACCTTGCTCAG  
CTTGGCTTCAGCTATATCAGTTCCTACTCCCTCAATCTTGCAACTCGTTTCATTAGATGAGATAGTAAATT  
GGACGGTGGAAATAGCCCTCAAGATAGTCCAAGAAATTTCTCCCCAGTGCCCTCAGCCCCATTTTTCATTTG  
CACGGAGATGCAATCCTTCACCTTCTTGGGCTCCTTCATCCCTCTCACATGCATAGCTTGTCTCTCTGTC  
CCAGGAAAACACCGAGAAAACAAAATGGCACCTTCAGCCAAGTTTCTGAATATTCAGGATAAACTCGGTT  
TATGGTCTGAATGTCTAGATATTTGTTTCAGTTTCTCAGAGACCTAACTATATCCTAAAGTGGATTAAA  
GAGAAGTTGCATCAGTTACCATAACCAACACACAGACGAGTTACACTTCTTATCAAAACATTTCTGTAC  
CACCGAAAGCATCGCCACTGAGAACAGATGCAGGAACACGCCGATGCGCCCCGTTCCCGAAGTCTGAGAC  
ATTGTATTTCTAAACCTATTAATTTAGAAGAGAAGGCACAGTTGTCTGTCTCTCCAGTTAGACTGGCTACA  
GCTCAGATGGAAGACGTCTAAAGGAAATATCACCAGCTACTCTCTGACAACGTCTACCCCTTAGCAGA  
TGGAGTGCTTAGTTTCACTCACCACCAGATTATTGAACTGGCTCGAGATTGCTTGATAAATCCCACCAGG  
GCCTCATCACCTCACGATACTTCTTGAATTACAGCACAATTAGATAAGTTGCTACAGGAGGCTCATGAT  
CGTTCAGAAAGTGGAGAATTGGCATTATTAACAACCTAGTTTCGAAAGATCCTAATTGTTATTGCCCGCCC  
TGCTCGGTTATTAGAGTGCTTGAATTTGATCCGGAAGAATTTTACTACCTATTGGAAGCAGCAGAAGGCC  
ATGCCAAAGAAGGACAGGGTATTAAAACCGACATTTCCAGGTACATCATTAGCCAACTGGGACTCAATAAG  
GATCCCTTGGAAGAAATGGCTCATTGGGAACTACGATAGTGGGACAGCAGAAACACCAGAAACAGATGA  
ATCAGTGAGTAGCTCTAATGCCTCCCTGAACTTCGAAGGAAACCTCGGGAAAGTGATTTTGAAACGATTA  
AATTGATTAGCAATGGAGCCTATGGGGCAGTCTACTTTGTTTCGGCATAAAGAATCCCGGCAGAGGTTTGGC  
ATGAAGAGATTAATAAACAGAACCTCATCTTCGAAGACAGATCCAGCAGGCCTTTGTGGAGCGGGATAT  
CTGACTTTTGCAGAAAACCCCTTTGTTGTGCTAGCATGATTGCTCCTTTGAAACAAGGCGCCACTTGTGCA  
TGGTCATGGAATATGTGGAAGGGGGAGACTGTGCTACTTTAATGAAAAACATGGGTCTCTCCCTGTTGAT  
ATGGCCAGAATGTACTTTGCTGAGACGGTCTTGGCCTTGAATATTTACATAATTATGGAATTGTACACAG  
GGATTTGAAACCAGACAACCTTGTGGTTACCTCCATGGGGCACATAAAGCTGACAGATTTTGGATTATCTA  
AGGTGGGACTAATGAGCATGACTACCAACCTTTACGAGGGTCATATTGAGAAGGATGCTAGAGAGTTCTTG  
GATAAACAGGTCTGTGGCACACCTGAATACATTGCACCAGAAGTGATTCTGAGGCAGGGTTATGGAAGCC  
GGTGGACTGGTGGGCCATGGGGATTATCCTCTATGAATTTCTGTTTGGATGCGTGCCATTCTTTGGGGATA  
CTCCAGAGGAGCTATTTGGACAAGTCATCAGTGATGAGATCAACTGGCCTGAGAAGGATGAGGCACCCCCA  
CCTGATGCCCAGGATCTGATTACCTTACTCCTCAGGCAGAATCCCCTGGAGAGGCTGGGAACAGGTGGTG  
ATATGAAGTCAAACAGCATCGATTCTTCCGTTCTTTAGACTGGAACAGTTTGCTGAGACAGAAGGCAGAAT  
TTATTTCCCAACTGGAATCTGAGGATGACACAAGTTATTTTGATACTCGGTCTGAGAAGTATCATCATATG  
GAAACGGAGGAAGAAGATGACACAAATGATGAAGACTTTAATGTGGAATAAGGCAGTTTCTTTCATGTTT  
ACACAGGTTTTCAAAAGTTTTCAGCAGTATAGATCGAATCACTCAGAATTCAGCAGAAGAGAAGGAAGACT  
CTGTGGACAAAACCAAAGCACCACCTTGCCATCCACAGAAACACTGAGCTGGAGTTTCAATATTCTGAA  
ATGCAACAGCTATCAACATCCAACCTTTCAGATACTGAAAGCAACAGACATAAACTCAGTTCTGGCCTACT  
TCCCAAATGGCTATTTCAACAGAGGGAGAGCAAGATGAAGCTGCCTCCTGCCCCTGGAGACCCCCATGAGG  
AGCCAGGAAGCCAGCCCTTCTCCTGAAGAGTGTGCCAGGAGGAGCCTGAGGCTCACCACCCAGCCAGC  
ACCATCAGCAGCTCCACCTGTCTAGTTGGCAGTTTTCAGAGCACTTGGATCAGATAAATGGACGAAGCGA  
GTGTGTGGACAGTACAGATAATTCCTCAAAGCCATCCAGTGAACCCGCTTCTCACATGGCTCGGCAGCGAT  
TAGAAAGCACAGAAAAAAGAAAAATCTCGGGGAAAGTCACAAAGTCCCTCTCTGCCAGTGCTCTTTCCCTC  
ATGATCCAGGAGATATGTTTGTGTTTCCCTCTGGGAAGTCCAATGTCTCCCATTCCTGTCTCGGA

Figure 1C

CCCTTCTTCTTCACGAGATTCTCTCTCCAGCCGAGATTCTCTCAGCAGCTTCTGCCAGTCCACATCAGCCGA  
TTGTGATCCACAGTTCGGGGAAGAATACTACGGCTTTACCATCCGAGCCATCCGGGTGTATGTGGGAGACAGT  
GACATCTATACAGTGCACCATATCGTCTGGAAATGTAGAAGAAGGAAGTCCGGCATGCCAGGCAGGACTGAA  
GGCTGGAGATCTTATCACTCACATCAATGGAGAACCAGTGCATGGACTTGTCCACACAGAAGTTATAGAAC  
TCCTACTGAAGAGTGGGAATAAGGTGTCAATCACTACTACCCCATTTGAAAAACACATCAATCAAACTGGA  
CCAGCCAGGAGAAACAGCTATAAGAGCCGGATGGTGAGGCGGAGCAAGAAATCCAAGAAGAAAGAAAGTCT  
CGAAAGGAGGAGATCTCTTTTCAAAAAGCTAGCCAAGCAGCCTTCTCCTTTACTCCACAGCCGGAAGTT  
TCTCCTGCTGAACAGATCCCTGTCTCGGGTGAGAGCCTCCAGGTTCCCCCACTCATAGCTTGTCTCCC  
CGGTCTCCAACACCAAGCTACCGCTCCACCCCTGACTTCCCATCTGGTACTAATTCTCTCCAGAGCAGCTC  
CCCTAGTTCTAGTGCCCCCAATTCCCCAGCAGGGTCCGGGCACATCCGGCCAGCACTCTCCACGGTCTTG  
CACCCAACTCGGCGGGCAGCGGTACCGGTCCGGAAGGCGAAAGTCCGCGGCAACATCCCACTGTCCCCG  
CTGGCCCCGACGCCCTCTCCAACCCGCAACCCACCTCCCCGAGCGGTACCATCCCCCTCTTCTGGGACA  
CTCACTGGGCAATTCCAAGATCGCGCAAGCCTTTCCAGCAAGATGCACTCCCGCCCACTCATCGTCAGAC  
ACATCGTGAGGCCCAAGAGTGCAGGAGCCCCCAGGTCCCCGCTGCTCAAGCGCGTGAGTCCGAGGAGAAG  
CTGTGCGCCTCTTACGGCAGTGACAAGAAGCACCTGTGCTCCCGCAAGCACAGCCTGGAGGTGACCCAAGA  
GGAGGTGCAGCGGGAGCAGTCCAGCGGGAGGCGCCGCTGCAGAGCCTGGATGAGAAGTGTGCGACGTGC  
CGCCGCTCAGCCGCGCCCGGCCAGTGGAGCAAGGCTGCCTGAAACGCCCAGTCTCCCGGAAGGTGGGCCGC  
CAGGAGTCTGTGGACGACCTGGACCGGACAAGCTGAAGGCCAAGGTGGTGGTGAAGAAAGCAGACGGCTT  
CCAGAGAAACAGGAATCCCAACAGAAATCCCATGGACCCGGGAGTGATTGGAAAACCTTTGCTCTGTTTA  
AGCTGGAAGAGAGAGAGAAGAAAGTCTATCCGAAGGCTGTGGAAAGGTCAAGTACTTTTGAAGCAAAAGCG  
TCTATGCAGGAGGCGCCACCGCTGGGCAGCCTGTGTAAGGATGCTCTTACAAGCAGGCCAGCGTGCAGCG  
CAGCGAGGGTGCATGTGCGATGGCCCCGGTGCCTGCGGAGCACCGCCAGGGTGGCGGGGACTTCAGACGGG  
CCCCCGCTCTTGGCACCTCCAGGATGGTCTCTGCCATCTCCCTCGACAGGGGCATCTCTGGGAAGGGGGAA  
GGCACGGAGAAGTCTCCAGGCCAAGGAGCTTCTCCGATGTGAAAAGTTAGACAGCAAGCTGGCCAACAT  
CGATTACCTCCGAAAGAAAATGTCACTTGAGGACAAAGAGGACAACCTCTGCCCTGTGCTGAAGCCCAAGA  
TGACAGCTGGCTCCCAAGATGCCTGCCAGGGAACCCAGTCCGACCCAGGGTGGGCAGCAGGAGCCCCCG  
CCGGCTTCTGAGAGCCGAGCTTTTGTGAGCAGCACCCATGCAGCTCAGATGAGTGCCGTCTCTTTTGTTC  
CCTCAAGGCCTTAACAGGCCGGGTGGACAGTGGAAACGAGAAGCCTGGCTTGGTTGCTCTGAGTCCCCTG  
TTAGGAAGAGCCCCCTCCGAGTATAAGCTGGAAGGTAGGTCTGTCTCATGCCCTGGAGCCGATCGAGGGCACT  
CTGGACATTGCTCTCTGTCCGGACCTCAGGCCTCCAAGACAGAACTGCCTTCCCCAGAGTCTGCACAGAG  
CCCCAGCCCAAGTGGTGAGTGAGGGCCTCTGTGCCACCAGTTCTCCCCAGCAGCAGTGGGAAAAAGAACG  
ATACCACAGTGCAAGAGAGCTTTCTCTCTCCAGCTTAAAGATGAATAAATCTTACCTGCTGGAGCCTTGG  
TTCTTGCCCCCAGCCGAGGTCTCCAGAATTCAACAGCAGTTTCCCTGCCTGACCCAGAGTTCAAGAGGGA  
CAGGAAAGGTCCCCATCTTACTGCCAGGAGCCCTGGAACAGTCATGGAAGCAATCCCCAACAGAGAGAGG  
GCAGCTCCCCCTAAACACCAAGACCACCACTGACCCCAAGCTTCTGACCTGCCTGGGGCAGAACCTCCAC  
AGCCCTGACCTGGCCAGGCCAGCTGCCCGCTCCACCTGAAGCTTCCCCCTCAAGGGAGAAGCCAGGCCT  
GAGGGAATCGTCTGAAAGAGGCCCTCCACAGCCAGAAGCGAGCGCTCTGCTGCGAGGGCTGACACATGCA  
GAGAGCCCTCCATGGAAGTGTGCTTTCCAGAACTGCGAAAACAGTGACAACTCCAAAAATCTCTCTCT  
GTGGGAAGGACCCACCCAGATTTCTATACACAGACCCAGGCCATGGAGAAAGCATGGGCGCCGGGTGGGAA  
AACGAACCACAAAGATGGCCAGGTGAGGCGAGGCCCGCCGAGAGCAACTCTCTGCACTCAGCTG  
GAATTCCTGTGAGAAGGAGCTGGGCAAGGTGAGGCGTGGCGTGAACCCAAGCCCGAAGCGCTTCTTGCC  
AGGCGGTCTCTGCAGCCACCTGGAATTGAGAGTGAGAAGAGTGAAAAGCTCTCCAGTTTCCCATCTTTGCA  
GAAAGATGGTGCCAAGGAACCTGAAAGGAAGGAGCAGCTCTACAAAGGCATCCAGCAGCATCCCTCCGC  
CCCCCTGACGGCCAAAGACCTGTCCAGCCCGGCTGCCAGGCAGCATTGCAGTTCCCCAAGCCACGCTTCT  
GGCAGAGAGCCGGGGGCAAGCCCAAGCACTGCAGAGCCAGCTCGAGCCCCAGGACCCCTCCCAAGCCTGT  
TGCTGCGCACAGTGAAAGCAGCAGCCCAAGCCCCGGCCTGGCCCTGACCCGGGCCCCCTCCAAAGACTAAGC  
ACCCCGACCGGTCCCTCTCTCTCAGAAAACCAAGTGTCCGGGGCCACAAAGGGCAAGAGCCTGCCACTCAG  
TCCCTCGGTGGCTCTAGCAGAGAGGGGAAGGGCCACAGTAAGAGTGGGCGGATGTGTTTCTCTGTACCCCC  
ACAGGCTCCTCTAGACGCCAAGCCACAACCCACAGTGGTGGGCGGCCCTGGAGGTGCTGGAGAAGCCT  
GTGCATTTGGCCAAGGCCGGACACCCAGGGCCTAGTGAGCCAGCGGACCAGAACTGTCCGCTGTTGGTGA  
AAAGCAAAACCTGTCTCCAAAGCACCCCAACCATCCACTGTGAAAGATTGCCCCACCTGTGCAACAGA  
CAGACAACAGACAGACAGACAAAAGCCCCGAGTCAGCCGGCCCAACACCGACAGAAGGGCGGAAGGGAAG

Figure 1D



AAATGCACTGAAGCACTTTATGCTCCAGCAGAGGGCGACAAGCTCGAGGCCGGCCTTTCCCTTTGTGCATAG  
CGAGAACCGGTTGAAAGGCGCGGAGCGGCCAGCCGCGGGGGTGGGGAAGGGCTTCCCTGAGGCCAGAGGGA  
AAGGGCCCGGTCCCAGAAAGCCACCGACGGAGGCAGACAAGCCCAATGGCATGAAACGGTCCCCCTCAGCC  
ACTGGGCAGAGTTCTTTCCGATCCACGGCCCTCCCGGAAAAGTCTCTGAGCTGCTCCTCCAGCTTCCCTGA  
AACCAGGGCCGGAGTTAGAGAGGCCTCTGCAGCCAGCAGCGACACCTCTTCTGCCAAGGCCGCCGGGGGCA  
TGCTGGAGCTTCCAGCCCCCAGCAACAGGGACCATAGGAAGGCTCAGCCTGCCGGGGAGGGCCGAACCCAC  
ATGACAAAGAGTGACTCCCTGCCCTCCTTCCGGGTCTCCACCTGCCTCTGGAGTCAACCAACCCCGACCC  
AAACACCATGGGCGGGGCCAGCCACCGGGACAGGGCTCTCTCGGTGACTGCCACCGTAGGGGAAACCAAAG  
GGAAGGACCCTGCCCCAGCCCAGCCTCCCCAGCTAGGAAACAGAACGTGGGCAGAGACGTGACCAAGCCA  
TCCCCAGCCCCAAACACTGACCGCCCCATCTCTCTTTCTAATGAGAAGGACTTTGTGGTACGGCAGAGGCG  
GGGGAAGAGAGTTTGCCTAGCAGCCCTCACAAAAGGCCCTTGTAAACGGGGAGGGCCAGGGGCAGGACTG  
TGGAGACCCGTCTCTGAACGGGCGACTGTGTCTTGACTACCTTTCAAAACCAGCACTGTGTGGGAATGTCCG  
CCAGGCAGAGCTCGGAGCCTCATTGAGACAGGGGAGAGAGAAAAGACAAAGAGGGGACCTTCTCCAGATGC  
CTTCCCAGTTGTAACCGGTAAACTGTTACCAGATAGTGTTTGTAC

>SGK021\_ID#NA\_4

ATGGGAGCGAACACTTCAAGAAAACCACCAGTGTTTTGATGAAAATGAAGATGTCAACTTTGACCACTTTGA  
AATTTTGCAGGCCATTGGGAAAGGCAGTTTTGGGAAGGTCTGCATTGTACAGAAGAATGATACCAAGAAGA  
TGTAACGAATGAAGTACATGAATAAAACAAAAGTGCCTGGAGCGCAATGAAGTGAGAAATGTCTTCAAGGAA  
CTCCAGATCATGCAGGGTCTGGAGCACCTTTCTGGTTAATTTGTGGTATTCTTCCAAGATGAGGAAGA  
CATGTTTCATGGTGGTGGACCTCCTGCTGGGTGGAGACCTGCGTTATCACCTGCAACAGAACGTCCACTTCA  
AGGAAGAAACAGTGAAGCTCTTCATCTGTGAGCTGGTCATGGCCCTGGACTACCTGCAGAACAGCGCATC  
ATTCACAGGGATATGAAGCCTGACAATATTTTACTTGACGAACATGGGCACGTGCACATCACAGATTTCAA  
CATTGCTGCGATGCTGCCAGGGAGACAGATTACCACCATGGCTGGCACCAAGCCTTACATGGCACCTG  
AGATGTTTCAGCTCCAGAAAAGGAGCAGGCTATTCTTTGCTGTTGACTGGTGGTCCCTGGGAGTGACGGCA  
TATGAACTGCTGAGAGGCCGGAGACCGTATCATATTCGCTCCAGTACTTCCAGCAAGGAAATGTACACAC  
GTTTGAGACGACTGTTGTAACCTTACCCTTCTGCCTGGTCACAGGAAATGGTGTCACTTCTTAAAAAGGTAA  
GAAGGAAGACTGCATGTCCAAACGAAGTAACAAAAGGAAGCAGGCTCTCTGGCTTAAGTTTGAAGTTAGT  
ATACAATATTGGGGACACCAACTGTCTTCTTGCAGAAAGGCAGGCTGAATTGTGATCCTACCTTTGAACT  
TGAGGAAATGATTTTGGAGTCCAAACCTCTACATAAGAAAAAAGCGTCTGGCAAAG

>SGK410\_ID#NA\_5

ATCTTTGGGCTGCTTGGGCGAGGCGACCCCTTGGGTGCGCGTTGCGGGTGAGGTGGGCAGGTAGGTGGGCGG  
ATGGCCGCGGTTCTCCGACAGGCGCGGGCGGGGCTTCCACGGCGTCCGAAGCCCCCTCCCCGACCC  
CCGGCTCCAGCGTTGAGGCGGGGAGTGAGGAGATGCCGACCCAGAGGGACAGCAGCACCATGTCCACA  
GGGTTGCAGGCGGCTGCAGTGGGGACCATTTCCACCAGGTCTGGGTGAAAGCCTACTACCGCGGGGATATC  
ATGATAACACATTTTGAACCTTCTGTCTCCTTTGAGGGCCTTTACAATGTGGTTCGAGACATGTGTTCTTT  
TGACAACGAACAGCTCTTACCATGAAATGGATAGACGAGGAAGGAGACCCGTGTACAGTATCATCTCAGT  
TGGAGTTAGAAGAAGCCTTTACACTTTATGAGCTAAACAAGGACTCTGAACTCTTGATTATGTGTTCCCT  
TGTGTACCAGAACGTCCCGGATGCCTTGTCCAAGAGAAGATAAATCCATCTACCGTAGAGGTGCACGCCA  
CTGGAGAAAGCTTTATTGTGCCAATGGACACACTTTTCAAGCCAAGTGCTTCAACAGGCGTGTGACTGTG  
CCATCTGCACAGACAGAATATGGGGACTTGGATGCCAAGGATATAAGTGTATCAACTGCAAACTCTTGTT  
CATAAGAAGTGCCATAAACTCCTCACAATTGAATGTGGGCGGCATTCTTTGCCACTGGAACCAATGATGCC  
CATGGATCAGTCATCCATGCATTCTGACCATGCACAGACAGTAATTCCATATAATCCTTCAAGTCATGAGA  
GTTTGGATCAAGTTGGTGAAGAAAAAGAGGCAATGAACACCAGAGAAAGTGGCAAAGCTTCATCCAGTCTA  
GGTCTTCAGGATTTTGATTTGCTCCAGGTAATAGGAAGAGGAAGTTATGCCAAAGTACTGTTGGTTCGATT  
AAAAAAAACAGATCGTATTTATGCAATGAAAGTTGTGAAAAAGAGCTTGTAAATGATGATGAGGATATTG  
ATTGGGTGCAGACAGAGAAACATGTGTTTGGAGCAGGCATCCAATCATCCTTTCTTGTGTTGGGCTGCAGTCT  
TGCTTTTCAGACAGAAAGCAGATTGTTCTTTGTTATAGAGTATGTAAATGGAGGAGACCTAATGTTTCATAT  
GCAGCGACAAAGAAAACCTCCTGAAGAACATGCCAGATTTTACTCTGCAGAAATCAGTCTAGCATTCAATT  
ATCTTCATGAGAGAGGGATAATTTATAGAGATTTGAAACTGGACAATGTATTACTGGACTCTGAAGGCCAC  
ATTAAACTCACTGACTACGGCATGTGTAAGGAAGGATTACGGCCAGGAGATACAACCAGCACTTTCTGTGG  
TACTCCTAATTACATTGCTCCTGAAATTTAAGAGGAGGAGATTATGGTTTCAGTGTGACTGGTGGGCTC

Figure 1E

TTGGAGTGCTAATGTTTGGAGACGATGGCAGGAAGGACTCCATTTGATACTGTTGGGAGCTCCGATAACCCCT  
GACCAGAACACAGAGGATTCTCTCTTCCAAGTTATTGTGGAAAAACAAATTCACATACCAGTTCTCTGTCTC  
TGGAAAAGCTGCAAGTGTCTGAAGAGTTTCTTAACAAGGACCCTAAGAAACGACTGGGTTGTCTATCCCTC  
AAACAGGATTTGCTGATATTAGGGACACCCGTTCTTCCAAAATGTTGATTGGGATATGATGGAGCAAAAA  
CAGGTGGTATCTCCCTTTAAACCAAGTATTTCTGGGGAATTTGGTTTGGACAACCTTTGATTCTCAGTTTAC  
TAATGAACCTGTCTGGCTCACTCCAGATGTCGATGACATTGTGAGAAAGATTGATCAGTCTGAATTTGAAG  
GTTTTGAGTATATCAATCCTCTTTTGATGTCTGCAGAAGAATGTGTCTGATCCTCATTTTTTCAACCATGTA  
TTCTACTCATGTTGCCATTTAATGCATGGATAAACTTGCTGCCAGCCTGGATACAATTAACCATTTTTATAT  
TTGCCACCTACAAAAAACACCCAGTATCTTCTCTCGTAGACTACATGAATCAATTATTACATCTTTTTTA  
CTATGAAAAATAAATTAATACTAGCTTCCAGACAATCATGTCAAAATTTAGTTGAACTGGTTTTTCAGTTT  
TTAAAGGCCTACAGATGAGTAATGAAGTTATCTTTTTTGTTTAAAAAAAACAAAGA

>SGK069\_ID#NA\_6

CCTTCCCAGTTCTGCTCCACAGAGCACCACGGATTGTGATTGTCCCAGCTCCGTCCCCCTTTCTCAGACC  
CGTCAATCCTAAGAGCCAGGCCTGGTACCCATGGGTCTCCTTCCCTACCAAGGCCAGCCCGCGCCCCGAG  
TCCCATGGCACCTGCCAGACAGGACACGGAGCCTGGCTTGTGGAGGTTACGCCTGCAGGACGGGGACGAA  
CTAGATCCAAATCCAAGACCCAGCACCCAGCGGAGGTTCCAGAACTGGAACTGTGTTTCTAAGCAGGA  
AGAGATGCCCCGCAACAGTCTGAGGAAGGGCCGCGGAGGCGAGGGGCTTCGGAGGACAGCGAGGAGGAGG  
GTCTGGGCGGCTTGACATTAGAGGAGCTCCAGCAGGGCCAGGAGGCTGCCCGCGCGCTGGAGGACATGATG  
ACGCTGAGTGCTCAGACCCTGGTCCGAGCCGAGGTGACGAGCTCTACGAGGAAGTCCGTCCCCCTGGGCCA  
GGGTGCTATGGCCGCGTCTTCTGGTCAACCATCGTCAGAAAGGCACACCCCTGGCACTGAAGCAGCTCC  
CGAAACCCCGCAGTCCCTCCGTGGCTTCTGTACGAGTTCTGTGTGGGGCTCTCGCTGGGCGCGCACTCA  
GCCATCGTGACGGCCTACGGCATTGGCATCGAGTCGGCACACTCCTACAGCTTCCGTGACGGAGCCCGTCTC  
GCACGGGGACCTCATGGCCTTCATCCAGCCCAAGTGGGCTCCCGCAGCCCGCGGTGCAACCGCTGCGCCG  
CCCAGCTGGCCTCCGCGCTGGAGTACATCCAGCCCGCGGCTGGTGTACCGGGACCTGAAGCCGGAGAAC  
GTCCTGGTGTGCGACCCGGCTGCCGGCGCTTCAAGCTGACCGACTTCGGCCACACGAGGCTCGCGGGAC  
GCTGTGCGCCTGGCCGGGCGCCCATCCCCTACACGGCCCCCGAGCTCTGCGCGCCCCCGCGCTCCCCG  
AGGGCCTGCCCATTACAGCCCGCCTGGACGCTGGGCGCTGGGCGTCTGCTCTTCTGCTCCTCACGGGC  
TACTTCCCCTGGGACCGGCCCTGGCCGAGGCGGACCCCTTCTACGAGGACTTCTCATCTGGCAGGCGTC  
GGGCCAGCCCCGGGACCGCCCTCAGCCCTGGTTCCGCTGGCCCCCGCGGCCAGCGCTTCTGCGGGGGC  
TGCTGGACCCCTACCCCCGAAGGAGGAGCGCTGTGATCGCCATCAGGGAGCCTTGGGGCGCCCCCTGGAGG  
CAGCGGGAGGGCGAGGCGGAGGAGTGGGAGCGGTGGAAGAGGAGGCTGGGCAGGCGCCTGCTGTGGCCAC  
CGAGGTGTTGCTTGGAGCGGGCTGACAGCAGGTGTGCTCCTGTGGAGTGACCCCAAGGAGCCTTCGAATC  
CAGGTTCAAGCCATAGAGTTGGGGCTCTGACCTGGACAATCCGGGATGTCTGCTGCGCTCAGCGGGAG  
GAGGGGCTTGGGGGACGGGGGTGGGGAGCAGCCGCTCATCTATCAGGACCCCGAGCGTTATCCGCAC  
CCCCCAGCCCCCGCCCAAGTAAGTGGACCCCTGTTGGCCTTTGCGAGGACTGCGGGGCGCCAGCCTTCAT  
CACCTTCCACGACCTCTAAGCCGAGAGAGGAGGAGGAGGAGGAAATGGTTTCTGCTGAGGCTGCAGGAC  
CCAGAATCAGAG

>SGK110\_ID#NA\_7

ATGACCCCCACTCGGGGCTGGGGGCTCCTGGCCTGCTCCACAGACAGGGGCGAGGCGGGGCTCTGTGGC  
AGTGACCCCTAACCTCTCAGGACACAAAAACAGGTGGCCCCGAGTCCGGGAGGACACAGCCACAGCCCTCC  
AACGGCTGGTGGAGCTGACGACCAGCAGGGTGACCCCGGTGAGGAGCCTTCGGGACAGTACCACCTCATC  
CGGAAGCTGGGCTCCGGCTCCTACGGCCCGGTGCTCCTCGCCAGCCTACACAGGGGGGTCCAGCTGTGGC  
TCTGTGACACACCAGGCTGCTGCAGACCCTGGCAGGACCCCTACAGACCCCCCGCTATTTTGCCTTCGCC  
CAGGAGTACGCGCCCTGTGGGGACCTCAGCGGGATGCTGCAGGAAAGGGGCTCCCAAGAACTGCTGGTGAA  
GCGGGTGGTGGCCAGTTGGCAGGAGCTCTGGACTTCTCCACAGCCGGGGGCTGGTCCACGCAGATGTCA  
AACCTGACAACGTGCTGGTCTTCGACCCAGTCTGACGCGGTGTGGCCCTGGGAGACCTGGGTCTGACCCGG  
CCAGAGGGCAGCCCGACCCCGCCCAAGTGGCTCTGCCACGGCACCGCCTGAGCTCTGTCTCTGCTGCT  
ACCGCCCGACACCTGCTCTGCGGCCAGCCGTGGACTCTGGGGCCTGGGGGTGCTTCTCTCTGTGCTG  
CCACTGCTGTTTCCCTTGGGACGTGGCACTGGCCCCCAACCCCTGAGTTGAGGCGCTTCGCTGGCTGGGTG  
ACCACCAAGCCTCAGCCACCTCAGCCACCACCACCTGGGACCACTTGGCGCCCCAGCCCTGGCCTTGCT

Figure 1F

CCAGGGGCTTCTGGACCTGGATCCCGAGACTAGGAGCCCCCACTGGCTGTCTGGACTTCCTGGGGGACG  
ACTGGGGGTTGCAAGGGAACAGAGAGGGACCTGGGGTTTTGGGGAGCGCCGTGTCTATGAGGACAGGGAG  
GAGGGAGGCTCAAGCCTGGAGGAGTGGACAGATGAGGGTGATGACAGCAAAAGT

>SGK053\_CKLIK\_ID#NA\_8

ATGGCCCCGGGAGAACGGCGAGAGCAGCTCCTCCTGGAAAAAGCAAGCTGAAGACATCAAGAAGATCTTCGA  
GTTCAAAGAGACCCCTCGGAACCGGGGCCCTTTTCCGAAGTGGTTTTAGCTGAAGAGAAGGCAACTGGCAAGC  
TCTTTGCTGTGAAGTGATCCCTAAGAAGCGCTGAAGGGCAAGGAAAGCAGCATAGAGAATGAGATAGCC  
GTCCTGAGAAAGATTAAGCATGAAAATATTGTTGCCCTGGAAGACATTTATGAAAGCCCAATCACCTGTA  
CTTGGTCATGCAGCTGGTGTCCGGTGGAGAGCTGTTTGACCGGATAGTGGAGAAGGGGTTTTATACAGAGA  
AGGATGCCAGCACTCTGATCCGCCAAGTCTTGAGCGCCGTGTACTATCTCCACAGAATGGGCATCGTCCAC  
AGAGACCTCAAGCCCGAAAATCTCTTGTAAGTACAGTCAAGATGAGGAGTCCAAAATAATGATCAGTGACTT  
TGGATTGTCAAAAATGGAGGGCAAAGGAGATGTGATGTCCACTGCCCTGTGGAATCCAGGCTATGTGCTC  
CTGAAGTCTCGCCAGAAACCTTACAGCAAAGCCGTTGACTGCTGGTCCATCGGAGTGATTGCCTACATC  
TTGCTCTGCGGCTACCCCTCCTTTTTATGATGAAAATGACTCCAAGCTCTTTGAGCAGATCCTCAAGGCGGA  
ATATGAGTTTGACTCTCCCTACTGGGATGACATCTCCGACTCTGCAAAAGACTTCATTGGAACCTGATGG  
AGAAGGACCCGAATAAAAGATACACGTGTGAGCAGGCAGCTCGGCACCCATGGATCGCTGGTGACACAGCC  
CTCAACAAAAACATCCACGAGTCCGTGAGCGCCAGATCCGGAAAACTTTGCCAAGAGCAAATGGAGACA  
AGCATTTAATGCCACGGCCGTGCTGAGACATATGAGAAAATACACCTCGGCAGCAGCCTGGACAGTTCAA  
ATGCAAGTGTTCGAGCAGCCTCAGTTTGCCAGCCAAAAGACTGTGCGTATGTAGCAAAACCAGAATCC  
CTCAGCTGA

>SGK124\_ID#NA\_9

ATGCGAGCCACCCCTCTGGCTGCTCCTGCGGGTTCCCTGTCCAGGAAGAAGCGGTTGGAGTTGGATGACAA  
CTTAGATACCGAGCGTCCCGTCCAGAAACGAGCTCGAAGTGGGGCCCCAGCCAGACTGCCCCCTGCCTGT  
TGCCCCCTGAGCCCACTACTGCTCCAGATCGTGCAACTGCTGTGGCCACTGCCTCCCGTCTTGGGCCCTAT  
GTCCTCCTGGAGCCCGAGGAGGGCGGGCGGGCCTACCAGGCCCTGCCTGACCTACAGGCACTGAGTATAC  
CTGCAAGGTGTACCCCGTCCAGGAAGCCCTGGCCGTGCTGGAGCCCTATGCGCGGCTGCCCCCGACAAGC  
ATGTGGCTCGGCCCACTGAGGTCTGGCTGGTACCCAGCTCCTCTACGCCTTTTTCACTCGGACCCATGGG  
GACATGCACAGCCTGGTGCGAAGCCGCCACCGTATCCCTGAGCCTGAGGCTGCCGTGCTCTTCCGCCAGAT  
GGCCACCGCCCTGGCGCACTGTCAACAGCACGGTCTGGTCTGCGTGATCTCAAGCTGTGTGCTTTGTCT  
TCGCTGACCGTGAGAGGAAGAAGCTGGTGTGGAGAACCTGGAGGACTCCTGCGTGCTGACTGGGCCAGAT  
GATTCCCTGTGGGACAAGCACGCGTGCCAGCCTACGTGGGACCTGAGATACCTAGCTCACGGGCCCTCATA  
CTCGGGCAAGGCAGCCGATGTCTGGAGCCTGGGCGTGCGCTCTTACCATGCTGGCCGGCCACTACCCCT  
TCCAGGACTCGGAGCCTGTCTGCTCTTCGGCAAGATCCGCCGCGGGGCCCTACGCCCTTGCTGCAGGCCTC  
TCGGCCCCCTGCCCGTGTCTGTTCTGCTGCTCCTTTCGTGGGAGCCAGCTGAACGGCTCACAGCCACAGG  
CATCCTCCTGCACCCCTGGCTGCGACAGGACCCGATGCCCTTAGCCCCAACCCGATCCCATCTCTGGGAGG  
CTGCCCAGGTGGTCCCTGATGGACTGGGGCTGGACGAAGCCAGGGAAGAGGAGGGAGACAGAGAAGTGTT  
CTGTATGGCTAG

>SGK254\_CAMKA\_ID#NA\_10

ATGGAGGGGGGTCCAGCTGTCTGCTGCCAGGATCCTCGGGCAGAGCTGGTAGAACGGGTGGCAGCCATCGA  
TGTGACTCACTTGGAGGAGGCAGATGGTGGCCAGAGCCTACTAGAAACGGTGTGGACCCCCACCACGGG  
CCAGAGCTGCCTCTGTGATCCCTGGCAGTACTTCAAGACTGCTCCAGCCCGGCCCTAGCCTCTCAGCCAGG  
AAGCTTTCCCTACAGGAGCGGCCAGCAGGAAGCTATCTGGAGGCGCAGGCTGGGCCCTATGCCACGGGGCC  
TGCCAGCCACATCTCCCCCGGGCCTGGCGGAGGCCACCATCGAGTCCACACGTGGCCATCTCAGATG  
CAGAGGACTGCGTGACGTGAACAGTACAAGCTGCAGAGTGAGATTGGCAAGGTGGGGCTGACTGATGCC  
TATCTGCAGGGTGCCCTACGGTGTGGTGGGCTGGCCTACAACGAAAGTGAAGACAGACTATGCAATGAA  
AGTCTTTTCCAAAAAGAAAGTTACTGAAGCAGTATGGCTTTCCACGTGCGCCCTCCCCGAGAGGGTCCCAGG  
CTGCCCAGGGAGGACCAGCCAAGCAGCTGCTGCCCCCTGGAGCGGGTGTACCAGGAGATTGCCATCCTGAAG  
AAGCTGGACCACGTGAATGTGGTCAAACCTGATCGAGGTCCTGGATGACCCAGCTGAGGACAACCTCTATTT  
GGTTGACCTCCTGAGAAAGGGGCCCGTCAATGGAAGTGCCCTGTGACAAGCCCTTCTCGGAGGAGCAAGCTC  
GCCTCTACCTGCGGGACGTCATCCTGGGCCCTCGAGTACTTGCACTGCCAGAAGATCGTCCACAGGGACATC

Figure 1G

AAGCCATCCAACCTGCTCCTGGGGGATGATGGGCACGTGAAGATCGCCGACTTTGGCGTCAGCAACCAGTT  
TGAGGGGAACGACGCTCAGCTGTCCAGCACGGCGGGAACCCAGCATTTCATGGCCCCGAGGCCATTTCTG  
ATTCCGGCCAGAGCTTCAGTGGGAAGGCCCTTGATGTATGGGCCACTGGCGTCACGTTGTACTGCTTTGTC  
TATGGGAAGTGCCATTTCATCGACGATTTTCATCCTGGCCCTCCACAGGAAGATCAAGAATGAGCCCGTGGT  
GTTTCCTGAGGAGCCAGAAATCAGCGAGGAGCTCAAGGACCTGATCCTGAAGATGTTAGACAAGAATCCCC  
AGACGAGAATTGGGGTGCCAGACATCAAGTTGCACCCTTGGGTGACCAAGAACGGGGAGGAGCCCTTCCT  
TCGGAGGAGGAGCACTGCAGCGTGGTGGAGGTGACAGAGGAGGAGTTAAGAACTCAGTCAGGCTCATCCC  
CAGCTGGACCACGGTGTATCCTGGTGAAGTCCATGCTGAGGAAGCGTTCCTTTGGGAACCCGTTTGAGCCCC  
AAGCACGGAGGGAAGAGCGATCCATGTCTGCTCCAGGAAACCTACTGGTGAAAGAAGGGTTTGGTGAAGGG  
GGCAAGAGCCCAGAGCTCCCCGGCTCCAGGAAGACGAGGCTGCATCCTGA

>SGK297\_CAMKIB2\_ID#NA\_11

ATGCTGTGCTGAAGAAACACCGGAGGACATCAGCAGCGTCTACGAGATCCGCGAGAGGCTCGGCTCGGG  
TGCCTTCTCCGAGGTGGTGTGGCCAGGAGCGGGGCTCCGCACACCTCGTGGCCCTCAAGTGCATCCCCA  
AGAAGGCCCTCCGGGGCAAGGAGGCCCTGGTGGAGAACGAGATCGCAGTGTCCGTAGGATCAGTCACCCC  
AACATCGTCGCTCTGGAGGATGTCCACGAGAGCCCTTCCACCTCTACCTGGCCATGGAACCTGGTGACGGG  
TGGCGAGCTGTTTGACCGCATCATGGAGCGCGGCTCCTACACAGAGAAGGATGCCAGCCATCTGGTGGGTC  
AGGTCCTTGGCGCCGTCTCCTACCTGCACAGCCTGGGATGTGTCACCGGGACCTCAAGCCCCAAAACCTC  
CTGTATGCCACGCCCTTTGAGGACTCGAAGATCATGGTCTCTGACTTTGGACTCTCCAAAATCCAGGCTGG  
GAACATGCTAGGCACCGCTGTGGGACCCCTGGATATGTGGCCCCAGAGCTCTTGGAGCAGAAACCTACG  
GGAAGGCCGTAGATGTGTGGGCCCTGGGCGTCATCTCTACATCCTGTGTGGGTACCCCCCTTCTAC  
GACGAGAGCGACCCTGAGCTCTTCAGCCAGATCCTGAGGGCCAGCTATGAGTTGACTCTCTTTCTGGGG  
TGACATCTCAGAATCAGCCAAAGACTTCATCCGGCACCTTCTGGAGCGAGACCCCCAGAAGAGGTTCACT  
GCCAACAGGCCCTTGGCGCACCTTTGGGTCTCTGGGGACACAGCCTTCGACAGGGACATCTTAGGCTCTGTC  
AGTGAGCAGATCCGGAAGACTTTGCTCGGACACACTGGAAGCGAGCCTTCAATGCCACCTCGTTCTGCG  
CCACATCCGGAAGCTGGGGCAGATCCCAGAGGGCGAGGGGCTCTGAGCAGGGCATGGCCCCCACAGCC  
ACTCAGGCCTCCGTGCTGGCCAGCCCCCAAGTGGTGA

>SGK411\_CAMKII DELTA2\_ID#NA\_12

ATGGCTTCGACCACCACCTGCACCAGGTTACGGACGAGTATCAGCTTTTCGAGGAGCTTGGAAGGGGGC  
ATTCTCAGTGGTGAGAAGATGTATGAAAATTCCTACTGGACAAGGATATGCTGCCAAAATTATCAACACCA  
AAAAGCTTTCTGCTAGGGATCATCAGAACTAGAAAAGAGAAGCTAGAATCTGCCGTCTTTTGAAGCACCT  
AATATTGTGCGACTTCATGATAGCATATCAGAAGAGGGCTTTCACTACTTGGTGTGTTGATTTAGTTACTGG  
AGGTGAACGTGTTTGAAGACATAGTGGCAAGAGAATACTACAGTGAAGCTGATGCCAGTCATTGTATACAGC  
AGATTCTAGAAAAGTGTAAATCATTGTACCTAAATGGCATAAGTTACAGGGACCTGAAGCCTGAGAATTTG  
CTTTTAGCTAGCAAATCCAAGGGAGCAGCTGTGAAATTGGCAGACTTTGGCTTAGCCATAGAAGTTCAAGG  
GGACCAGCAGGCGTGGTTTGGTTTTGCTGGCACACCTGGATATCTTCTCCAGAAGTTTACGTAAAGATC  
CTTATGGAAGCCAGTGGATATGTGGGCATGTGGTGTATTCTCTATATCTACTTGTGGGGTATCCACCC  
TTCTGGGATGAAGACCAACACAGACTCTATCAGCAGATCAAGGCTGGAGCTTATGATTTTCCATCACCAGA  
ATGGGACACGGTGACTCCTGAAGCCAAAGACCTCATCAATAAAATGCTTACTATCAACCCTGCCAAACGCA  
TCACAGCCTCAGAGGCACTGAAGCACCCATGGATCTGTCAACGTTCTACTGTTGCTTCCATGATGCACAGA  
CAGGAGACTGTAGACTGCTTGAAGAAATTTAATGCTAGAAGAAAATAAGGGTGCCATCTTGACAACTAT  
GCTGGCTACAAGGAATTTCTCAGCAGCCAAGAGTTTGTGTAAGAAACCAGATGGAGTAAAGGAGTCAACTG  
AGAGTTCAAATACAACAATTGAGGATGAAGATGTGAAAGCACGAAAGCAAGAGATTATCAAAGTCACTGAA  
CAACTGATCGAAGCTATCAACAATGGGGACTTTGAAGCCTACACAAAATCTGTGACCCAGGCCTTACTGC  
TTTGAACCTGAAGCTTTGGGTAAATTTAGTGGAAGGGATGGATTTTCACCGATTCTACTTTGAAAATGCTT  
TGTCAAAAGCAATAAACCAATCCACACTATTATTCTAAACCCTCATGTACATCTGGTAGGGGATGATGCC  
GCCTGCATAGCATATATTAGGCTCACACAGTACATGGATGGCAGTGAATGCCAAAGACAATGCAGTCAGA  
AGAGACTCGTGTGTGGCACCGCCGGGATGGAAGTGGCAGAATGTTTATTTTCATCGCTCGGGGTACCAA  
CAGTACCCATCAAGCCACCCTGTATTCCAAATGGGAAGAAAATCTCTCAGGAGGCACCTCTTGTGGCAA  
AACATCTGA

>SGK027\_ID#NA\_13

Figure 1H

ATGACTGCAGTGTATATGAATGGAGGTGGCCTGGTGAACCCCCACTATGCCCGGTGGGATCGGCGCGACAG  
TGTAGAAAGTGGCTGTCAGACCGAGAGTAGCAAGGAGGGTGAGGAGGGACAGCCCCGCCAGCTGACGCCCT  
TCGAGAAACTGACACAGGACATGTCCAGGATGAGAAAGTGGTGAGGGAGATCACGCTGGGGAACGGATA  
GGCTTCTACCGAATTCGAGGGGAAATCGGAAGTGGAACTTCTCCCAAGTGAAGCTTGGGATTCACTCCCT  
AACCAGAAAGAGGTGGCCATTAAGATCCTGGACAAGACCAAGTTAGACCAGAAAACCCAGAGGCTACTAT  
CCCCAGAAATCTCCAGCATGGAAAAGCTGCACCATCCCAACATCATCCGCCCTTACGAAGTGGTGGAGACC  
CTATCCAAGCTGCACTTGGTGATGGAGTATGCAGGGGGTGGGGAGCTCTTCGGAAAAATTAGCACTGAGGG  
GAAGCTCTCTGAACCAGAAAGCAAGCTCATCTTCTCCAGATTGTGTCTGCCGTGAAGCACATGCATGAAA  
ACCAAATTATTATAGAGATCTGAAAGCAGAAAATGTATTCTATACCAGTAATACTTGTGTGAAGGTGGGC  
GATTTTGGATTACAGCACAGTAAGCAAAAAGGTGAAATGCTGAACACTTTCTGTGGGTCTCCTCCCTACGC  
TGCGCCTGAACTCTTCCGGGACGAGCACTACATCGGCATTTACGTGGATATCTGGGCCCTTGGGGGTGCTTT  
TGTAATTCATGGTGACTGGCACCATGCCATTTTCGGGCAGAAACCGTGGCCAACTAAAAAGAGCATCCTC  
GAGGGCACATACAGTGTAACCGCCGACGTGTGAGAGCCCTGCCACCGACTCATCCGAGGAGTCTTTCAGCA  
GATCCCCACGGAGAGGTACGGAATCGACTGCATCATGAATGATGAATGGATGCAAGGGGTGCCATACCCTA  
CACCTTTGGAACTTTTCCAACCTGGATCCCAACATTTGTTCGGAAACAGCACTCTCAAGGAAGAAGAAAT  
GAGGTCAAAAAGCACTTTAGAACATTTGGGCATTACAGAAGAGCATATTGCAAAATACCAAGGGAGAGATGC  
TCGCAGCTCAATCACAGGGGTCTATAGAATTATTTTACATAGAGTCCAAAGGAAGAAGGCTTTGGAAAGTG  
TCCAGTCATGATGCTACCAGACCCTAAAGAAAGAGACCTCAAAAAAGGGTCCCGTGTCTACAGAGGGATA  
AGACACACATCCAAATTTTGCTCGATTTTATAA

>SGK046B\_ID#NA\_14

GAAGAGGAGGAGGCCAGACCATGTTTGGGCAAATCTGTAGCAATGTAGTACTGCCTCTATAAAACGT  
CGCACATGCAGACGTAAGCTTACAAATCATCTACTCCCTGAAGAGCGCAATATAAAATCGTGGACTTTG  
GCTTTAGCATGACATTCGGGGAAAGGAGATTTCGGAACCCCTTTTGTGGCATGTACCCCTATGTGGCCACA  
GAACTCTTCCTG

>SGK046C\_ID#NA\_15

TTCAGGCAGATTCTGTCAAGCCTGCAGTACTGCCGCTTAAAGAACATCACACATCCAGAGCTAAAGCCACA  
AACCATCTAGTCATAGAGGAAGGCAATATAAAAAATAACGGATTTT

>SGK089\_ID#NA\_16

AATATCTTCTTTGACAAAGAACGTAACATCAAAATAGCCAACTATGGGTTTAGTTCAACATTTGCAAGAG  
AGACAACACTGAAAAGGGAAACGTTTGGTGGGACTCTCCCTTATGCAGGCCAGAACTCTTCTTGGGCCATG  
GGTACCAATGCCCTGCCATCAACGTTTGGAGCCTTGGCATCATCGTATGCAAAATGGTGGCTGGGGCCCTG  
CCATTTTATGAGCAGAACTGTAAGGATATACTGAAAAAA

>SGK133\_ID#NA\_17

ATGGTGGCGGGTTAACTTTGGGGAAGGGCCCGAGTCCCCGGATGGTGATGTGAGCGTGCCGGAGAGAAA  
GGACGAGGTGGCGGGGGGAGGCGGAGAGGAGGAGGCGGGAAGAGAGAGGGCGCCACGCCCAATATGTGG  
GCCCTATCGGCTGGAGAAGACGCTGGGCAAAGGACAGACAGGGCTGGTTAACTCGGGGTCCACTGCATC  
ACGGGTCAGAAGGTCGCCATCAAGATCGTGAACCGGGAGAAGCTGTGCGAGTGGTGCTGATGAAGGTGGA  
GCGGGAGATCGCCATCCTGAAGCTCATCGAACCCACATGTCTCAAGCTCCACGACGTCTACGAGAACA  
AGAAATATTTGTACCTGGTTCTGGAGCACGTCTCGGGGGGTGAGCTATTGACTACCTGGTAAAGAAGGGG  
AGACTGACGCCCAAGGAGGCCCGAAAGTTCTTCCGCCAGATTGTGTCTGCGCTGGACTTCTGCCACAGCTA  
CTCCATCTGCCACAGAGACCTAAAGCCCGAGAACCTGCTTTTGGATGAGAAAAACAACATCCGCATTGCAG  
ACTTCGGCATGGCGTCCCTGCAGGTGGGGGACAGCCTCCTGGAGACCAGCTGCGGGTCCCCCATATGCG  
TGTCAGAGGTGATTAAGGGGAAAAATATGATGGCCGCCGGGCAGACATGTGGAGCTGTGGAGTCACTCT  
CTTCGCCCTGCTCGTGGGGGCTCTGCCCTTTGATGACGACAACCTCCGCCAGCTGTGGAGAAGGTGAAAC  
GGGGCGTCTTCCACATGCCCCCACTTATCTCCAGATTGCCAGAGCCTCCTGAGGGGAATGATCGAAGTG  
GAGCCCGAAAAAAGGCTCAGTCTGGAGCAAATTCAGAAACATCCTTGGTACCTAGGCGGGAAACACGAGCC  
AGACCCGTGCCTGGAGCCAGCCCTGGCCGCCGGGTAGCCATGCGGAGCCTGCCATCCAACGGAGAGCTGG  
ACCCCGACGTCTTAGAGAGCATGGCATCACTGGGCTGCTTCAGGGACCGGAGAGGCTGCATCGCGAGCTG  
CGCAGTGAGGAGGAGAACCAAGAAAAGATGATATATTATCTGCTTTTGGATCGGAAGGAGCGGTATCCAG

Figure 11

CTGTGAGGACCAGGACCTGCCTCCCCGGAATGATGTTGACCCCCCGGAAGCGTGTGGATTCTCCCATGC  
 TGAGCCGTCACGGGAAGCGGCGACCAGAGCGGAAGTCCATGGAAGTCCTGAGCATCACCGATGCCGGGGGT  
 GGTGGCTCCCTGTACCCACCCGACGGGCTTGGAGATGGCCAGCACAGCCAGAGATCCCGTAGCGTCAG  
 TGGAGCCTCCACGGGTCTGTCTCCAGCCCTCTAAGCAGCCCAAGGAGTCCGGTCTTTTCTTTTACC  
 AGCCGGGGGTGGAGATGAGGCTCGAGGCGGGGCTCCCCGACTTCCAAAACGACAGCTGCCTTCTCGG  
 GGCCCCAGGGTGGGGCGCCGGGAGCAGCCCCCGCCCCCAGTGGCCGCTCCACACCCCTGCCCGGCC  
 CCCAGGCTCCCCGCGCTCCTCTGGCGGACCCCTTGCACTCGCCTCTGCACACGCCCCGGGCCAGTCCCA  
 CCGGGACCCCGGGGACAACACCACCCCCAGCCCCGGCGGTGGCGTCGGGGGAGCCGCTGGAGGAGTCGT  
 CTAACCTCCATCCGCAACAGCTTCCCTGGGCTCCCCCTCGCTTTCACCGGCGCAAGATGCAGGTCCCTACCGC  
 TGAGGAGATGTCCAGCTTGACGCCAGAGTCTCCCCGGAGCTGGCAAACGCTCCTGGTTCCGGAACTTCA  
 TCTCCTTGACAAAAGAAGAAATATTCTCGTGCTAAAGGACAAACCTCTCAGCAGCATCAAAGCAGAC  
 ATCGTCCATGCCTTTCTGTGATCCCCAGCCTGAGTCACAGTGTGCTGTGCACAGACCAGCTTCAGGGCCGA  
 GTACAAGGCCAGTGGCGGCCCTCCGTCTTCCAAAAGCCCGTCCGCTTCAGGTGGACATCAGTCTCTCTG  
 AGGTTCCAGAGCCCTCCCCGCGACGGGACGGCAGCGGAGGTGGTGGCATCTACTCCGTACCTTCACTCTC  
 ATCTCGGTTCCCGAGCCCTGGCGAGTGGTGGAGACCATCCAGGCACAGCTCCTGAGCACTCATGA  
 CCAGCCCTCCGTGCAGGCCCTGGCGAGCAGAGAAGAACGGGGCCAGACCCGGCCTGCTGGTGCCCCACCC  
 GAAGCTGCAGCCCCACCCGGCCGCCAGACCCAGAGCTGAGCAGCTCTCCCCGCGAGGCCCCCCCAAG  
 GACAAGAAGCTCCTGGCCACCAACGGGACCCCTCTGCCCTGA

>SGK004\_MSK\_ID#NA\_18

ATGGTTATCATGTGCGAGTTCAGCGCGGACCCCGGGGCCAGGGTCAGGGCCAGCAGAAGCCCTCCGGGT  
 GGGTTTTTACGACATCGAGCGGACCTGGGCAAAGGCAACTTCGCGGTGGTGAAGCTGGCGGGCATCGAG  
 TCACAAAACGCAGGTTGCAATAAAAATAATTGATAAAACACGATTAGATTCAAGCAATTTGGAGAAAATC  
 TATCGTGAGGTTTCTGATGAAGCTTCTGAACCATCCACACATCATAAAGCTTTACCAGGTTATGAAAC  
 AAAGGACATGCTTTACATCGTCACTGAATTTGCTAAAAATGGAGAAATGTTTGATTATTTGACTTCCAACG  
 GGCACCTGAGTGAGAACGAGGCGCGGAAGAAGTCTGGCAAATCCTGTGCGCCGTGGAGTACTGTACGAC  
 CATCACATCGTCCACCGGGACCTCAAGACCGAGAACCCTCCTGTGGATGGCAACATGGACATCAAGCTGGC  
 AGGCACGGAGGATTTTGGATTTGGGAATTTCTACAAGTCAGGAGAGCCTCTGTCCACGTGGTGTGGAGCC  
 CCCCCTATGCGCCCCCGGAAGTCTTTGAGGGGAAGGAGTATGAAGGCCCCAGCTGGACATCTGGAGCCTG  
 GCGTGGTCTGTACGTCTGTGCTGCGGTTCTCTCCCTTCGATGGGCCTAACCTGCCGACGCTGAGACA  
 GCGGGTGTGGAGGGCGCTTCCGCATCCCCCTTCTCATGTCTCAAGACTGTGAGAGCCTGATCCGCCGCA  
 TGCTGGTGGTGGACCCCGCCAGGCGCATCAACCTAGCCAGATCCGGCAGCACCCTGGATGCGGGCTGAG  
 CCTGTCTGCGGGACCCGCTGCCCCGCTTCTCCGACACAGCTACACCTCAACCTGGGCGACTACGA  
 TGAGCAGGCGCTGGGTATCATGCAGACCTGGGCGTGGACCCGAGAGGACGTTGAGTCAAGCAAAACA  
 GCAGCTATAACCACTTTGCTGCCATTTATTACCTCCTCCTTGAGCGGCTCAAGGAGTATCCGAATGCCAG  
 TGCGCCCCCGGGCCTGCCAGGCAGCCGCGGCTCGGAGCTCGGACCTCAGTGGTTTGGAGGTGCCCTCA  
 GGAAGGTCTTTCCACCGACCTTTCCGACCTGCCTTGCTGTGCGCGCAGCCGACAGCTTGGTGCAGTCCG  
 TCCTCCAGGCCGAGATGGACTGTGAGCTCCAGAGCTCGCTGCAGTGGCCCTTGTCTTCCCGGTGGATGCC  
 AGCTGCAGCGGAGTGTTCGGCCCCCGGCCGTGTCCTCAAGCAGCCTGCTGGACACAGCCATCAGTGAGGA  
 GGCCAGGCAGGGGCGGGCCTAGAGGAGGAGCAGGACACGCAGGAGTCCCTGCCCAGCAGCAGGGCCGGA  
 GGCACACCTTGGCCGAGGTCTCCACCCGCTTCTCCCCACTCACCGCGCATGTATAGTCTCTCCCCCTCC  
 ACCACGGCAAGTCTGCAGAGGGAACAGCTCTGACAGTTGTCTGACCTTCTCTGCGAGCAAAAGCCCCGC  
 GGGGCTCAGTGGCACCCCGGCCACTCAGGGGCTGCTGGGCGCCTGCTCCCCGCTCAGGCTGGCCTCGCCCT  
 TCCTGGGGTTCGAGTCCGCCACCCAGTGTGCTGACGGCTCAGGGGGGCTTGGGAGGAGCTGTTCTGCTCCCT  
 GTCAGCTTCCAGGAGGACGGCGGGCGTCCGACACCTCACTGACTCAAGGGCTGAAGGCCTTTCCGGCAGCA  
 GCTGAGGAAGACACGCGGACCAAAGGTTTCTGGGACTGAACAAATCAAGGGGCTGGCTCGCCAGGTGT  
 GCCAGGCCCGCCAGCCGGGAGGAGGGCTGAGCCCTTCCACGCCCCCTGCACAGAGCCAGGC  
 CTGCACGGCGGCGCAGCCGGCAGCCGGGAGGGCTGGAGCTGCTGGAGGAGGTGCTAGAGCAGCAGAGGCT  
 GCTCCAGTTACAGCACCACCCGGCCGCTGCACCCGGCTGCTCCAGGCCCGCCAGCCCGGCTGCCCCGT  
 TTGTGATCGCCCCCTGTGATGGCCCTGGGGCTGCCCCGCTCCCCAGCACCTCTCAGCTCGGGGCTCCCCG  
 CTGCTGCCGCCCCACTCCTGCAGACCGCGCGTCCCCGGTGGCCTCAGCGGCGCAGCTCCTGGACACACA  
 CCTGCACATTGGCACCGGCCCCACCGCCCTCCCCGCTGTGCCCCCACCACGCCTGGCCAGGCTGGCCCCAG

Figure 1J

GTTGTGAGCCCCCTGGGGCTGCTGCAGGGGGACTGTGAGATGGAGGACCTGATGCCCTGCTCCCTAGGCACG  
TTTGTCTTGGTGCAGTGA

>SGK006\_ID#NA\_19

GTGCCTCTGTATTTTCAGGAAGCTCTCCCATGATCTGCAAATGACTTTCCCTTACACTAAGAAAGTTGGGGC  
TTATTTGGTTGGCAAATGATCAACAAGGGTCCTTTTGCCAAAGTCATAGAAGGGCTGGGTCTCCCCATGG  
GTGAAAAGGCAACCATCAAAGGCATTAGCAAGAAGAAAGCCAAAGCAAGACTCCTATGTTCTAAAGAATATA  
AAGTGTGAATCTTGGATCCAGTAAATTGTCAAATACCCTAATGCTTCCCAGCTACAAACCCTGGAGACTAA  
GAACCTCTTATTATGTGATGATGGAGCTATGCCTCAGTGGCAACCTCATGAACAGGATCTGGGATCAAAAGA  
AACTAAGTGAAGGGAAGTGAAGATGATCTTCAGGCAGATGTTGTCCGCAATGGAGCATGTGCACTGGCAT  
GGAATGGTACACAAGGATTATGCATTTTTTTTCTTCTCAGGCAGGAAAAGCCACCAAGCCCCCAAGTTTA  
TGAAACCCCTTGACATTAATACTGAGAAGTTGGAGGGTCTTCTCAGGAGCAGTTAAGTACTCAGGGTGGAA  
GTCTTGCCATATGCGGCCCTGAACTGCTTGCTCATCAGAAATATGGCCCCAAAGTGGATGTCTGGTTCATG  
GGTAAGTTTCTGATGCTGACTGGCACCTTACCGTTCACATTGGAGATATTAGTATTGAGCAACTCTATCA  
GAAGATGGTGAATTGGTGCAATCAGCAATGTCTTACCTGAAATCTCCTTTGGTAATTTACTGTGGTCTCTAA  
GTTTGGTA

>SGK180\_SNRK\_ID#NA\_20

ATGGCAGGATTTAAGCGAGGGTATGATGGAAAGATTGCTGGATTATATGATCTGGATAAAACCTTGGGTGCG  
AGGCCATTTTGCCTGTGGTACTTGCTCGAGGCATGTCTTTACGGGTGAAAAGGTGGCAGTAAAAGTTATTG  
ACAAGACAAAACCTGGACACTCTAGCTACTGGTCATCTTTCCAGGAAGTGAGATGCATGAACTAGTGCAG  
CATCCTAACATCGTCCGCCTTTATGAAGTTATTGACACCCAGACCAAACTATATCTTATTCTAGAAGTTGG  
GGATGGAGGAGATATGTTGATTATATAATGAAACATGAGGAGGGTCTTAATGAAGACTTGGCCAAGAAGT  
ATTTTGCTCAGATAGTTCATGCTATATCTTATTGCCATAAACTCCATGTGGTTCACAGAGACTTAAACCA  
GAGAATGTAGTCTTCTTTGAAAAACAAGTCTTTGTAAAGTTGACAGACTTTGGGTTTCAGCAACAAATTTCA  
ACCAGGGAAGAAGCTCACTACAAGCTGTGGATCTCTTGTCATATTCCGCTCCAGAAATCTGCTTGGTGTATG  
AGTATGATGCACCTGCAGTAGATATTTGGAGTCTGGGAGTGATCCTTTTTCATGTTGGTGTGTGGGCAGCCG  
CCCTTTCAAGAAGCCAATGACAGTGAAACACTGACAATGATCATGGATTGCAATATACAGTACCATCCCA  
TGTGTCTAAAGAGTGTAAGACCTAATCACACGGATGCTACAGAGAGATCCCAAGAGAAGGGCTTCTTTAG  
AAGAGATTGAAAATCATCCTTGGCTTCAGGGAGTGGACCCCTTACCAGCTACAAAGTATAACATTCCCTT  
GTGTACATACAAAAATCTCTCGGAAGAGGAGCACAACAGCATCATTACAGCGCATGGTGTCTGGGGACATAGC  
GGATCGAGACGCCATTGTAGAAGCCCTGGAAACCAACAGGTATAACCATATCACAGCCACATACTTCTGTC  
TGGCTGAAAGGATCCTGAGAGAAAAGCAAGAGAGAAAGAAATACAGACCAGATCTGCAAGCCCGAGCAATATC  
AAGGCCCAGTTTAGGCAGTCATGGCCAACCAAAATTGATGTACCCAGGACCTTGAGGATGACCTCACGGC  
CACTCCTTTGTCCACGCGACTGTCCCTCAGTCTCCTGCTCGGGCTGCTGACAGTGTCTCAATGGCCACA  
GGAGCAAAGGCCCTGTGTGACTCAGCTAAGAAAGATGACCTCCCTGAGTTGGCTGGACCAGCACTCTCTACG  
GTGCCACCCGCAAGCTTAAAACCCACAGCCAGTGGGCGGAAGTGTCTGTTACAGGTGGAAGAAGATGAAGA  
GGAAGATGAGGAGGACAAGAAACCCATGTCCCTCTCAACACAAGTGGTTTTTGCGCCGGAAGCCATCTGTAA  
CCATCCGCTGACATCCAGGAAGAGTGCGCCCTCTCAACCAGATCTTTGAGGAAGGGGAATCTGACGAT  
GAGTTTGACATGGATGAGAATCTGCCTCCCAAGTTGAGCAGGTTAAAGATGAATATAGCTTCTCCAGGTAC  
AGTTCAAAAACGCTACCACCGGAGGAAAAGTCAGGGCCGGGGCTCCAGCTGCAGTAGTTTCGGAGACCAGTG  
ATGATGATTCTGAAAGCCGGCGGGCTCGATAAAGATAGCGGGTTACCTACTCTTGGCACCAGCGGGAT  
AGCAGCGAGGGGGCCCCCTGGCAGTGAGGGGGATGGCGGGGGCCAGAGCAAGCCGAGCAATGCCAGTGGAGG  
GGTGGACAAGGCCAGCCCCAGTGAGAACAATGCTGGTGGGGGAGTCCCTCCAGCGGCTCGGGTGGCAACC  
CCACCAATACATCGGGTACCACAGCCGCTGTGCCGGCCCCAGCAACTCCATGCAGCTGGCCTCTCGCAGT  
GTGGGGAGCTCGTTGAGAGCCTCAAACCTCATGAGCCTCTGCCTCGGCTCCAGCTTCATGGGAGCACC  
GTACATTATTGATCCACAGAATGGCTTGTCAATTTCCAGTGTGAAAGTCCAAGAGAAATCTACGTGGAAAA  
TGTGCATTAGCTCCACAGGGAATGCAGGGCAGGTCCCTGCAGTGGGCGGCATAAAGTTTCTCTGACCAC  
ATGGCAGATACCACCACTGAATTGGAACGGATAAAGAGCAAGAACCTGAAAAATAACGTGCTGCAGCTACC  
TCTGTGCGAAAAAGACCATCTCTGTGAACATCCAGCGGAACCCCTAAGGAGGGGCTGCTGTGCGCATCCAGCC  
CAGCCAGCTGTTGCCATGTCATCTGA

>SGK386\_MLCKS\_ID#NA\_21

Figure 1K



ATGGCGACAGAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGGGAAGGAAAGGAGGGTGGGA  
TCCTGATGGGTGTCTCACCTCCTCTGCAGACAAGGCACCTAAAGGTCCACAGGTGAAAGACCCCTGGCTG  
CAGGGAAAGACCCCTGGCCCCCAGACCCAAAGAAAGCTCCGGATCCACCACCCCTGAAGAAAGATGCCAAA  
GCCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCCAACCTCAACTAGCAGCCAAGGCCCAAAGGAGA  
GGGTGACAGGGGCGGGGGGCGCGGAGGGCAGTGCTGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCCGA  
CACCTGAGACCAGCGTCAAGAAGCCCAAGGCTGAGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCC  
AGGGTGGGCAAGAAGGCAGCAGAGGGCCAAGCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCC  
CAGCTGTCTGCCATCATCTCCAGTTCTGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGC  
TCACCTTTGAAGGGGTGCCCATGACCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAG  
AACATCCTGGCAGAGAGCCAGAAGGAAGTGGGAGAGAAAACCCCAAGGCCAGGCTGGCCAGGCTAAGATGCA  
AGGGGACACCTCGAGGGGGATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCT  
GTCTCACAGCCAGGGAGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCT  
CACCGCATGGTGGAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAATCCAAGGAGGCGCTCGG  
AGGTGGCAAGTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCA  
TCAAGAAACAGACTCCCAAAGACAAGGAAATGGTGTGTCTGGAGATTGAGGTGATGAACCAAGCTGAACCA  
CGCAATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCTGTTTATGGAGTACATCGA  
GGGCGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTGTTG  
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTGCACCTGGACCTCAAGCCAGAG  
AACATCCTGTGTGTCAACACCACCGGGCATTGTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAGGTATAA  
CCCCAACGAGAAGCTGAAGGTGAACCTTTGGGACCCAGAGTTTCTGTACCTGAGGTGGTGAATTATGACC  
AAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTGAGCGGCTCTCCCCC  
TTCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTGGTACTTTGATGATGAGAC  
CTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCAACCTCATCGTCAAGGACCAGAGGGCCCGGA  
TGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAGAAAGCCAAACGCTGTAACCGA  
CGCCTTAAGTCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCGCTGGAAGAAAAACTTCATTGCTGT  
CAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCACTGATGGCTCTGGGGGT

>SGK003\_ID#NA\_22

ATGACAAACAACAGCGGCTCCAAAGCCGAACCTCGTTGTGGGAGGGAAATACAAACTGGTGCGGAAGATCGG  
GTCTGGCTCCTTTGGAGACGTTTATCTGGGCATCACCACCACCAACGGCGAGGACGTAGCAGTGAAGCTGG  
AATCTCAGAAGGTCAAGCACCCCAAGTTGCTGTATGAGAGCAAACTCTACACGATTCTTCAAGGTGGGGTT  
GGCATCCCCACATGCACTGGTATGGTCAGGAAAAAGACAACAATGTGCTAGTCATGGACCTTCTGGGACC  
CAGCCTCGAAGACCTCTTTAATTTCTGTTCAAGAAGGTTTACCATGAAACTGTACTTATGTTAGCCGACC  
AGATGATCAGCAGAATTGAATACGTGCATACAAAGAAATTTCTACACCGAGACATTAAACCAGATAACTTC  
CTGATGGGTACTGGGCGTCACTGTAATAAGTTGTTCTTATTGATTTTGGTCTGGCCAAAAAGTACAGAGA  
CAACAGGACCAGGCAACACATACCGTACAGAGAAGATAAAACACCTCATTGGCACTGTCCGATATGCCAGCA  
TCAATGCACATCTTGGTATTGAGCAGAGCGCCGAGATGACATGGAATCCTTAGGCTACGTTTTCTATGTAT  
TTTAATAGAACCAGCCTGCCGTGGCAAGGACTAAGGGCTATGACAAAAAACAATAATATGAAAAGATTAG  
TGAGAAGAAGATGTCCACCCCTGTTGAAGTTTATGTAAGGGGTTTCTGCGAGAATTGCCATGTACTTGA  
ACTACTGTGCTGGGCTGCGCTTTGAGGAAGTCCAGATTACATGTATCTGAGGCAGCTATTCCGCATTCTT  
TTCAGGACCCCTGAACCACCAATATGACTACACATTTGATTGGACGATGTTAAAGCAGAAAGCAGCACAGCA  
GGCAGCCTCTTCCAGTGGGCAGGGTCAGCAGGCCCAACCCAGACAGGCAAGCAAACTGAAAAAACAAGA  
ATAATGTGAAAGATAACTAA

>SGK066\_ID#NA\_23

ATGAGTGGGGGAGGAGAGCAGCCGGATATCCTGAGTGTGGAATCCTAGTGAAAGAAAGATGGAAAGTGTT  
GAGAAAGATTGGGGGTGGGGGCTTTGGAGAAATTTACGATGCTTGGACATGCTCACCAGGGAAATGTTG  
CACTGAAGGTGGAATCAGCTCAANNNCAAAAACAAGTTCTGAAAATGGAAGTTGCTGTTTTGAAAAAGCTG  
CAAGGTGGCTTTGTTCAAATTAGGATTTATATACAAGGCAGGCACATTGCATTATATTTGTTGGCCATGTC  
ACAGTTTGGTCCGAATCTGGCAGATCTTCGCGTAGCCAGTCCCGAGGCACATTACCATTTAGTACCACTC  
TCCGGCTGGGTAGACAGATTTTGGAGTCTATTGAAAGCATTCATTCTGTGGGATTCTTGCATCGAGACATC  
AAACCGTCGAACCTTCGCTATGGGTGCGCTTTCCTAGTACATGTAGGAAATGTTACATGCTTGATTTGGCTT  
GGCTCGACAATTTACCAATTCCTGTGGTGACGTGACACCACCTCGAGCTGTGGCAGGTTTTTCGAGGGACAG

Figure 1L



TTCGTTATGCATCAATCAACGCACATCGGAACAGGGAAATGGGAAGACATGATGACCTTTGGTCCTTATTC  
TACATGTTGGTGGAGTTTGTGGTTGGTCAGCTGCCCTGGAGAAAAATAAAGGACAAGGAGCAAGTAGGCTC  
TATTAAGGAGAGATATGACCACAGGCTCATGTTGAAACATCTCCCTCCAGAATTCAGCATCTTTCTAGACC  
ATATCTCTTCTTGGATTATTTTACAAAACCAGACTACCAGCTTCTTACATCCGTGTTTGACAATAGCATC  
AAGACTTTTGGAGTAATTGAGAGTGACCTTTTGACTGGGAGAAGACTGGAAATGATGGCTCCCTAACAAC  
CACCCTACTTCTACCACCCCTCAGTTGCACACTCGCTTGACCCCTGCTGCAATTGGAATTGCCAATGCTA  
CTCCCATCCCTGGAGACTTGCTTCGAGAAAAATACAGATGAGGTATTTCCAGATGAACAGCTTAGCGATGGA  
GAAATGGCATCCCTGTTGGTGTGTACCCAGATAAATTGCCTGGATCTCTGGGACACCCCGTCCCCAGGA  
GAAGGATGTTTGGGAAGAGATGGATGCCAACAAAAACAAGATAAAGCTTGAATTTGTAAGGTA

>SGK041\_NKIAMRE\_ID#NA\_24

ATGGAGATGTATGAAACCCTTGGAAAAGTGGGAGAGGGAAGTTACGGAACAGTCATGAAATGTAAACATAA  
GAATACTGGGCAGATAGTGGCCATTAAAGATATTTATGAGAGACCAGAACAATCTGTCAACAAAATTGCGA  
TGAGAGAAAATAAAGTTTCTAAAGCAATTTATCAGCAAAAACCTGGTCAATCTGATTGAAGTTTTAGACAG  
AAAAAGAAAATTCATTTGGTATTTGAATTTATTGACCACACAGTATTAGATGAGTTACAACATTATTGTCA  
TGGACTAGAGAGTAAGCGACTTAGAAAAACCTCTCCAGATCCCTCGAGCAATTGACTATCTTCACAGTA  
ATAATATCATTCATCGAGATATAAAACCTGAGAATATTTTAGTATCCAGTCAGGAATTACTAAGCTCTGT  
GATTTTGGTTTTGACGAACACTAGCAGCTCCTGGGGACATTTATACGGACTATGTGGCCACACGCTGGTA  
TAGAGCTCCCGAATTAGTATTAAGATACTTCTTATGGAAAACCTGTGGATATCTGGGCTTTGGGCTGTA  
TGATCATTGAGATGGCCACTGGAAATCCCTATCTTCTAGTAGTTCTGATTTGGATTTACTCCATAAAATT  
GTTTTGAAAGTGGGCAATTTGTACCTCACTTGACAGATATCTTTTCCAAGAGCCCCATTTTGTCTGGGGT  
AGTCTTCTCAAGTTCAACACCCCAAAAATGCAAGAAAAAATATCCAAAGCTTAATGGATTGTTGGCAG  
ATATAGTTTATGCTTTTACAAATTGATCCTGCTGACAGGATATCATCTAGTGATCTTTTGCATCATGAG  
TATTTTACTAGAGATGGATTTATTGAAAAATTCATGCCAGAAGTAAAGCTAAATTAATGTCAGGAAGCAAA  
AGTCAATTCATTAATAAAGCCAAAAGAGAGTTCTAAAGAAAATGAACTCAGGAAAGATGAAAGAAAAACAG  
TTTATACCAATACACTGCTAAGTAGTTAGTTTGGGAAAGGAAATAGAAAAAGAGAAAAAGCCCAAGGAG  
ATCAAAGTCAGAGTTATTAAAGTCAAAGGAGGAAGAGGAGATATCTCAGAACCAAAAAAGAAAGAGTATGA  
AGGTGGACTTGGTCAACAGGATGCAAATGAAAATGTTTATCTTATGTCTCCAGATACAAAACCTTGTAACCA  
TTGAACCACCAAAACCTATCAATCCCAGCACTAACTGTAATGGCTTGAAGAAAAATCCACATTGCGGAGGT  
TCTGTGACAAATGCCACCCATCAATCTAACTAACAGTAATTTGATGGCTGCAAATCTCAGTTCAAATCTCTT  
TCACCCAGTGTGAGGTAACTGAAAGAGCAAAAAGAGACGCACCTTCTTCACAATCTATTGGACAAGTTA  
TGCTTAATAGCAGGCAAGAGGATCCAGGTCTATTCAAAGCCAAATGGAGAAGGGTATATTTAATGAGCGA  
ACAGGTCAACAGTGACCAAAATGGCAAATGAGAACAAGGAAGCTGAATTTTCCAGATCTGACAGGAAAGA  
ATTCCATTTTCCAGAATTGCCTGTGACAAATACAGTCAAAAGATACAAAAGGAATGGAAGTTAAACAGATAA  
AAATGCTGAAGAGGGAGTCAAAGAAAACAGAGTCATCTAAGATACCAACTTTACTTAACGTGGATCAAAAT  
CAAGAAAAACAAGAGGGTGGAGATGGCCATTGCGAGGGGAAGAATTGAAGAGAAAACAGGTTTTTTTTTC

>SGK112\_ID#NA\_25

ATGGAAAAGTATGAAAAATTAGCTAAGACTGGAGAAGGGTCTTATGGGGTTGTATTCAAATGCAGAAACAA  
AACCTCTGGACAAGTAGTAGCTGTTAAAAAATTTGTGGAATCTGAAGATGATCCTGTTGTTAAGAAAATAG  
CACTAAGAGAAATACGTATGTTGAAGCAATTAACATCCAAATCTTGTGAACCTCATCGAGGTGTTTCAGG  
AGAAAAAGGAAAATGCATTTAGTTTTTGAATACTGTGATCATACACTTTTAAATGAGCTGGAAAGAAACCC  
AAATGGAGTTGCTGATGGAGTGATCAAAAGCGTATTATGGCAAACACTTCAAGCTCTTAATTTCTGTGATA  
TACATAACTGTATTACAGAGATATAAAACCTGAAAATATTTCTAATAACTAAGCAAGGAATAATCAAGATT  
TGTGACTTCGGGTTTGACAAATTTCTGAGTAATTCACATGTTGGGAGAATTGACTTACCTGATCTGATTGA  
TGCATTTGCACTCCAGGAGATGCCACCGATTATGTAGCTACGAGATGGTACCGAGCTCCTGAACTTC  
TTGTGGGAGATACTCAGTATGGTTCTTCAGTCGATATATGGGCTATTGGTTGTGTTTTGTCAGAGCTCCTG  
ACAGGCCAGCCACTGTGGCCTGGAAAATCAGATGTGGACCAACTTTATCTGATAATCAGAACTAGGAAA  
ATTAATCCCAAGACATCAATCAATCTTTAAAAGTAACGGGTTTTTCCATGGCATCAGTATACCTGAGCCG  
AAGACATGGAACTCTTGAGGAAAAGTTCTCAGATGTTTCATCCTGTGGCTCTGAACCTCATGAAGGGGTGT  
CTGAAGATGAATCCAGATGACAGATTAACCTGTTCCAACTCCTGGAGAGCTCCTACTTTGATTCTTTTCA  
AGAGGCCCAAATTAAGAAAAGCACGTAATGAAGGAAGAAACAGAAGACGCCAACAGAATCAACTGTTGCT

Figure 1M

CTCTCATACCAGGAAGCCACATCTCCCCACACCTGATGGAAGAAAACAAGTCTCCAGTTAAATTTGAT  
CACCTTCCAAACATTTAG

>SGK038\_ERK7\_ID#NA\_26

ATGTGCACCGTAGTGGACCCTCGCATTGTCCGGAGATACCTACTCAGGCGGCAGCTCGGGCAGGGGGCCTA  
TGGCATTGTGTGGAAGGCAGTGGACCGGAGGACTGGTGAGGTCGTGGCCATCAAGAAAATCTTTGATGCTT  
TTAGGGATAAGACAGATGCCAGGACATGGGCTTCCTTCTTGCTCCACCCACCCACACACCTGTGTTCTG  
TCTCTTCAGAGAACATTCCGGGAAATCACGCTCCTCCAGGAGTTTGGGGACCATCCCAACATCATCAGCCT  
CCTTGACGTGATCCGGGCAGAGAACGACAGGGACATTTACCTGGTGTGTTGAGTTTATGGACACTGACCTGA  
ACGCAGTCATCCGGAAGGGCGCCTGCTGCAGGACGTCCACGTGCGCTCCATCTTCTACCAGCTCCTGCGG  
GCCACCCGGTTCCTCCACTCGGGGCACGTTGTGCACCGGGACCAGAAGCCGTCCAATGTGCTCCTGGATGC  
CAACTGCACAGTGAAGCTGTGTGACTTTGGCCTGGCCCGCTCCCTGGGCGACCTCCCTGAGGGGCTGAGG  
ACCAGGCCGTGACAGAGTACGTGGCCACACGCTGGTACCGAGCACCAGGAGGTGCTGCTCTCTCGCACCGA  
TACACCTTTGGGTGGACATGTGGAGTCTGGGCTGTATCCTGGGGGAGATGCTGCGGGGAGACCCCTGTT  
CCCCGGCACGTCCACCCTCCACCAGCTGGAGCTGATCCTGGAGACCATCCCACCGCCATCTGAGGAGGACC  
TCCTGGCTCTCGGCTCAGGCTGCCGTGCTCTGTGCTGCACAGCTGGGGTCCCCGGCCACGACAGACGCTG  
GATGCCCTCCTACCGCCAGACACCTCCCCAGAGGCCCTTGGACCTCCTTAGGCGACTCCTGGTGTTCGCCCC  
GGACAAGCGGTTAAGCGCGACCCAGGCACTGCAGCACCCCTACGTGCAGAGGTTCCACTGCCCCAGCGACG  
AGTGGGCACGAGAGGCAGATGTGCGGCCCGGGCACACGAAGGGGTCCAGCTCTCTGTGCTGAGTACCGC  
AGCCGCGTCTATCAGATGATCCTGGAGTGTGGAGGCAGCAGCGGACCTCGAGAGAGAAGGGCCCGGAGGG  
TGTCTCCCCAAGCCAGGCACACCTGCACAAACCCAGAGCCGACCCCTCAGCTGCCCTTAGGACACCTGTGC  
AGGGTCCCAGACCCAGGCCCCAGAGCAGCCAGGCCATGACCCTGCCGAGCAGCCCAAGAACGTTCCCAGG  
CAGAACTCCGCTCCCCTGCTCCAACTGCTCTCCTAGGGAATGGGGAAAGGCCCTTGGGGCGAAGGAGC  
GCCCCCTTGACACTCTCGCTGGTGAAGCCAAGCGGGAGGGGAGCTGCGCCCTCCCTGACCTCCCAGGCTG  
CGGCTCAGGTGGCCAACCAGGCCCTGATCCGGGTGACTGGAACCGGGGCGGTGGGGTGAGGGTGGCCAGC  
GTACAACAGGTCCCTCCCCGGCTTCTCCGAGGCCCGGCCCGGAGGATGTTACGACCTCTGCCTT  
GCAGGGTGCCCAGGGGGTGCCAGGGCTTTGCTTGGAGGCTACTCCCAAGCCTACGGGACTGTCTGCCACT  
CGGCACTGGGCCACCTGCCCCTGCTGGAGGGGCACCATGTGTGA

>SGK158\_ID#NA\_27

ATGGCTGCCATATTGGGAGACACCATCATGGTGGCTAAAGGCCTTGTCAAGCTGACCCAGGCGGCCGTGGA  
AACCCACCTGCAGACTTGGGCATCGGAGGGGAGCTGATCATGGCAGCCAGGGCCCTGCAGTCCACGGCTG  
TGGAGCAGATTGGCATGTTCTTGGGAAGGTGCAGGGTCAGGATAAACATGAAGAATATTTTGCTGAGAAC  
TTCGGCGGCCCAGAAGGGGAGTTCCACTTCTCAGTCCCGCATGCAGCCGAGCCTCCACAGACTTCTCTTC  
AGCCTCCGCTCCCGACAGTCAGCGCCCCCATCCCTGGGTCATGCCACAGCGAGGGGCCAGCTCCTGCCT  
ACGTGGCCAGTGGACCCTTTAGAGAAGCCGGGTTCGCCGAGGCTTTCAGCGAAGGTTCTTCCACCAGGA  
GGGAGGCTCTTTGCAAACCCAGAGACTCATTCTCTGCCATGGGCTTTCAGCGAAGGTTCTTCCACCAGGA  
CCAATCCCCTGTTGGGGGCTCACAGCCGAGGACATTGAGAAGGCCCGCAGGCTAAGGCTCGCCCGGAGA  
ACAAGCAGCACAAACAGACGCTCAGCGAGCATGCCCGGAGCGGAAGGTGCCTGTGACGAGGATTGGCCGG  
CTGGCCAACTTCGGAGGTCTGGCCGTGGGCTTGGGCTTCGGGGCACTGGCAGAGGTGCGCAAGAAGAGCCT  
GCGCTCCGAGGACCCCTCAGGGAAGAAGGCCGTGCTGGGTTCAGTCCCTTCTGTCCGAGGCCAATGCAG  
AGCGGATCGTGCACGCTCTGCAAGGTGCGTGGTGGCGCACTCAAGCTGGGCCAGATGCTGAGCATCCAG  
GATGATGCCTTTATCAACCCCCACCCGGCGAGTGTCTACTCACGCCCGCACGACAGCCTGTGAGGCAGGG  
GCTGCCCAGACTGCTGGGGACGTGGAGGGTGCACTGTTGCCAGGTCTGGCCGTGGGCTTGGGCTTCGGGG  
CACTGGCAGAGGTGCGCAAGAAGAGCCTGCGCTCCGAGGACCCCTCAGGGAAGAAGGCCGTGCTGGGTTCC  
AGTCCTTTCTGTCCGAGGCAATGCAGAGCGGATCGTGCACGCTCTGCAAGGTGCGTGGTGGCGCACT  
CAAGCTGGGCCAGATGCTGAGCATCCAGGATGATGCCTTTATCAACCCCCACCTGGCTAAGATCTTCGAGC  
GGGTGCGGCAGAGCGCGGACTTCATGCCACTGAAGCAGATGATGAAAATCTCAACAACGACCTGGGCCCC  
AACTGGCGGGACAAGTTGGAATACTTCGAGGAGCGGCCCTTCGCCCGCGCATCCATTGGGCAGGTGCACTT  
GGCCCAATGAAGGGCGGCCGAGGTGGCCATGAAGATCCAGTACCCTGGCGTGGCCAGAGCATCAACA  
GTGATGTCAACAACCTCATGGCCGTGTTGAACATGAGCAACATGCTTCCAGAAGGCCGTTCCTCCGAGCAC  
CTGATCGACGTGCTGAGGCGGGAGCTGGCCCTGGAGTGTGACTACCAGCGAGAGGCCGCTGTGCCGCAA  
GTTACAGGGACCTGCTGAAGGGCCACCCCTTCTTCTATGTGCTGAGATTGTGGATGAGCTCTGCAGCCAC

Figure 1N

ATGTGCTGACCACAGAGCTGGTGTCTGGCTTCCCCCTGGACCAGGCCGAAGGGCTCAGCCAGGAGATTTCGG  
AACGAGATCTGCTACAACATCCTGGTTCTGTGCCTGAGGGAGCTGTTTCGAGTTCCACTTCATGCAAACAGA  
CCCCAACTGGTCCAACTTCTTCTATGACCCCCAGCAGCACAAAGGTGGCTCTTTTGGATTTTGGGGCAACGC  
GGGAATATGACAGATCCTTCACCGACCTCTACATTAGATCATCAGGGCTGCTGCCGACAGGGACAGGGAG  
ACTGTGCGGGGCGAAATCCATAGAGATGAAGTTCCCTACCGCTACGAGGTCAAGGTCATGGAAGACGCCCCA  
CTTGGATGCCATCCTCATCTGGGGGAGGCCTTCGCCTCTGATGAGCCTTTTGATTTTGGCACTCAGAGCA  
CCACCGAGAAGATCCACAACCTGATTCCCGTCATGCTGAGGCACCGTCTCGTCCCCCACCAGGAAACC  
TACTCCCTGCACAGGAAGATGGGGGGCTCCTTCCTCATCTGCTCCAAGCTGAAGGCCCGCTTCCCTTGCAA  
GGCCATGTTTCGAGGAGGCCTACAGCAACTACTGCAAGAGGCAGGCCAGCAGTAG

>SGK429\_ID#NA\_28

ATGGTGGCGCCCTGGCGCGTCTCCGTGAGGTTTGCCTGTGCGACCTGAGGTGCTTCGAGCTCAGACAGGG  
ACTCAGCCTCCTGAGGCCCTCCGAGTGCCCTCGCGATGCCAGGCTCTGCTGGCTTCGCTGGCACTTTGC  
CCAAGGTCGTCTCCTGTGCGGGGACGTGGGTGAGGGGGCCCCCTGACGTTCTGAGTCGGCGAAGGGTCCGC  
TGCAGCGGGGCGGCTGGCGCGGGGCCCCGCGAGAGCCTCCCCCGAGCGGACCTCTGGGCGGCGTCTTCCT  
GCATCTCCGCTCTGGCTTCGCGCGGGCGCTCTGTTGGTGAATTCCTCCCCCTCCTACTCCTCTACCCCC  
TCACCTACCTGGCTCCAGCGTCTCCACCCTCTGGCTCCACCTGCTTCTGAAAGCCACCAGACCTCAGGC  
CCAACCTACATCAAACCTGGGCCAGTGGGCCAGACCCGCGCGATCTGTTTTCGGAGGCTTTCTGTGCCCA  
ATTTTCCAAGCTGCATGTCCGAGTGACGCCCCACCCGTGGACTCACACTGAGCGCTTCCTTCGGCAGGCTT  
TTGGGGATGACTGGGGGAGCATCCTCTCTTTTGAACCCGGGAACCTGTGGGCTCAGGCTGCGTGGCCAG  
GTGTACAAAGCATACGCCAACACTGCCTTCCTGGAGACTGACAGCGTCCAGAGACTTGGCAGGGCCTCCTG  
TCTGCCGCCCTTCTACATACTGGGGCAGTCGGTGGGCTGAGAGAGCTCTTTGGATACCTTGGAATGGCC  
GGAAACCTCCAGAAAATCTCGCAGACCAGTCGTTTCTAGAAAGGCTGCTCCTCCCTAAAGCTGACCTGGTT  
GGATCAAATGCAGGGGTGTCTCGGGCTCAGGTCCCTGGCCACCAACCTGAGGCCACCAACCTCATCTCCGT  
GGCAGTGAAAGTGTTCACCCCTGGCCTGCTCGCTCAGGTGCATATGGACCTGCTGCTGATGAAGATTGGCA  
GCCGAGTCTCGGAGTTTTTGCCAGGCATCAAGTGGCTTAGCTTGCCCTGAGATTGTGGAGGAATTTGAGAAG  
CTGATGGTCCAACAGATTGACCTGCGTTACGAAGCTCAGAATCTAGAACTTCCAGGTCAACTTCCGGAA  
TGTGAAAGCCGTCAAGTTCCCCACCCCTCTGCGCCCTTTGTCAACAGAGAAGTCTTGGTGGAAACGTATG  
AAGAGAGTGTGCTGTGTCCAGTTACCAGCAGGCAGGAATTCCTGGACTTGAAAAGGAAGATTGCACGG  
CTGGGATCAACATGCTCCTGAAGATGATATTTGTGGATAACTTTGTCCATGCAGACCTTCACCCTGGAAA  
CATCCTGGTTTCAGGGTGCCAACGGGCTGTCTCGAGTCAAGGAGCGCAGCTGCAGCAGGCGGACATCTGTG  
ACACTCTGGTGGTGCCCGTGCCATCTTCCCTCTGCCCCGCTGCGACTGGTGCTGCTGGATGCTGGCATTTGTG  
GCGGAGCTGCAGGCCCCCTGACCTGAGGAATTTCCGGGCAGTTTTTCATGGCTGTGGTATGGGGCAGGGCCA  
GAGAGTGGCTGAGCTGATCCTGCATCATGCCGGGCCAGCGAGTGCAGGGACGTGGAGGGGTTCAAACCG  
AGATGGCCATGCTGGTGACCCAGGCCAGGAAGAACCACCATCACCTGGAGAAGCTTCATGTGTCCAGCCTT  
CTCTCTAGTGTCTTTAAGTTGTGATGACTCACAAGGTAAAGCTTGAGAGCAACTTTGCCCTCCATTGTGTT  
TGCCATCATGGTGTGGAGGGGCTTGCCGCTCACTGGACCCCAAACCTGGACATCCTGGAGGCAGCGAGGC  
CCTTCCTCCTCACGGGCCAGTGTGCCCCCGTGA

>SGK152\_SUDD\_ID#NA\_29

ATGGATCTGGTAGGAGTGGCATCGCCTGAGCCCGGACGGCAGCGGCTGGGGACCCAGCAAGTGTCCATG  
GGCTATTCCTCAAAATACAATATCTTGTTCTTTGGCTGATGTAATGAGTGAACAGCTGGCCAAAGAATTGC  
AGTTAGAAGAAGAGCTGCCGTTTTTCTGAAGTTGCTGTTGCTGAAGGACCATTTATTACTGGAGAAAAC  
ATTGATACTTCCAGTGACCTTATGCTGGCTCAGATGCTACAGATGGAATATGACAGAGAATATGATGCACA  
GCTTAGGCGTGAAGAAAAAAATTCATGGAGATAGCAAAGTTCCATTTCTTTGAAAATTATCGAAAAG  
TGCATCCTTATGAAGACAGCGATAGCTCTGAAGATGAGGTGACTGGCAGGATACTCGTGATGATCCCTAC  
AGACCAGAAAACCGGTTCCCACTCCTAAAAAGGGCTTTATTGGAAAAGGAAAAGATACACCACCAACA  
TGATGAAGTAGTATGTGGGAGAAAAGAACACAGCAAGAATGGAAAATTTTGACCTGAGTTTCAGGTAGGAG  
ATGGAATTGGAATGGATTTAAAACTATCAAACCATGTTTCAATGCTTTAAAAACAACATGCCTACTCAGAA  
GAACGTGGAAGTGCCCGCTACATGAGAAAAGGAGCATTCTACAGCAGAAAAGCAGTTGATCCTAAGAC  
ACGTTTACTTATGTATAAAATGGTCAACTCTGGAATGTTGGAGACAATCACTGGCTGTATTAGTACAGGAA  
AGGAGTCTGTTGCTTTTCATGCATATGGAGGGAGCATGGAGGATGAAAAGGAAGATAGTAAAGTTATACCT  
ACAGAATGTGCCATCAAGGTATTTAAACAACCCCTTAATGAATTTAAGAATCGTGACAAATATATTAAAGA

Figure 10

TGATTTTCAGGTTTAAAGATCGCTTCAGTAAACTAAATCCACGTAAGATCCACCGCATGTGGGCAGAAAAAG  
AAATGCACAATCTCGCAAGAATGCAGAGAGCTGGAAATCCTTGTCCAACAGTTGTACTACTGAAGAAACAC  
ATTTTAGTTATGTCTTTTATTGGCCATGATCAAGTTCCAGCCCTAAATTAAAAGAAGTAAAGCTCAATAG  
TGAAGAAATGAAAGAAGCCTACTATCAAACCTTTCATTGTGATGCGGCAGTTATATCATGAATGTACGCTTG  
TCCATGCTGACCTCAGTGAGTATAACATGCTGTGGCATGCTGGAAAGGTCTGGTTGATCGATGTCAGTCAG  
TCAGTAGAACCTACCCACCCTCACGGCCTGGAGTTCTTGTTCGGGACTGCAGGAATGTCTCGCAGTTTTT  
CCAGAAAGGAGGAGTCAAGGAAGCCCTTAGTGAACGAGAACTCTTCAATGCTGTTTTAGGCTTAAACATCA  
CAGCAGATAATGAAGCTGATTTTTTAGCTGAGATAGAAGCTTTGGAGAAAATGAATGAAGATCACGTTTACG  
AAGAATGGAAGGAAAGCTGCTTCATTTTTGAAAGATGATGGAGACCCACCCTACTATATGATGAATAG

>SGK077\_ID#NA\_30

ATGGCGGCTTCGCTCCCGGGACCTGGGAGCCGGCTTTTCCGCACATATGGGGCTGCGGACGGCAGGAGACA  
GCGGCGGCCCGGGCCGGGAAGCCGCGCAGTGGTTCCCGCCGCGAGGACCGGAGGCGTTTCTTCAACAGCAGCG  
GCAGCAGCGACGCCAGCATCGGCGACCCCTCGCAGTCCGACGATCCTGACGATCCCGACGACCCCGACTTC  
CCCGGCAGCCCGGTGAGGCGGCGGCGGAGGTGTCCCGGCGGCGAGTGCCCAAGGACCGGCCAGCCTGAC  
CGTGACCCCAAAGCGCTGGAAGCTGCGAGCTCGCCCAAGCCTAACCGTGACCCCAAGACGCTGGGGCTGC  
GAGCTCGGCCCCCGCAGAAGTGCAGCACACCCTGCGGCCGCTCCGACTTCCGCCCTTCCCCAGCCGCGAC  
TCCGGCCGCTCAGCCCGACCTCAGCGTGTGCGGCCAGCCAGGACCGGACGAGCTGGGCATCAGTGC  
CTCCCTGTTTACGCTCTCTGGCCTCGCCCTGCCCCGGTCCCCAACGCCAAGGGACAGTGTCTCATCTCGATCG  
GCACCTCCGCTGTCTGGTTGCAGCCTCAGCCGTCCGAGCGACCTCCACCTCCAGAAAGTCTCCCTGGAC  
CGAGCATCTCTCCCTGCTCCAGGAGGAAGCGACAGGAGGAGCAAGGACACAGGATGGTCCACCAAAC  
CCGCGCCAGCCTCAGGTGAGTTCTCTTTGGCCTTATGAACTCAGGAACCCCTGAGGATTTCTGAGTTTCGGG  
CAGATGGGAAGAATATGAGAGAGTCTGTGTAAAGGAAACTGGTGGTGGGAAATGACCCAGAGGGTCCCA  
GGTCTGTCAAGCACAGGCAAGAGGAGGGCCACAGGCCAGGACTCTGTGCAAGAGAGAGGGCTTCAAGAGGC  
CGTCCCGAGAGAGCATCAGGAGGCCAGTGTTCCTCAAGGGCCGATTGTGCCAAGGGGAACAGACAGGCTGG  
AGAGAACTAGATCAAGCCGGAAGAGCAAACATCAGGAGGCAACGGAACCTCTCTCTCCATTCCCACCGC  
TTTAAAAAGGGCCAAAAGCTGGGAAAAGATTGTTCCCCACCCAGGACCTGACTCCTTTACAGAATGCCTG  
CTTTTGGACCAAAACCAGGGCTTCCTTCAGTTTCCACAAGAAAGAAATTGTGACTGATGTGTGAGAGGTCT  
GCAGCATCTATACCACTGCCACTTCTCTCTCTGGATCCCTCCTATCAGAATGTTCAAACCGGCCTGTCTATG  
AACAGAACAAAGTGGTGCTCCGTCCTCTTGGCACTCCTCCTCTATGTATTTGCTAAGCCCCCTTAAACACTCT  
AAGTATTTCAAACAAAAGGCATCTGATGCTGAAAAGGTTTATGGGGAATGCAGTCAGAAGGGTCTCTGTCC  
CCTTTAGCCATTGCTTCCACAGAAAACTGCAACGCTGTGAGAAGATTGGGGAAGGGGTGTTTGGCGAA  
GTGTTTCAAACAATTGCTGATCACACCCGCTAGCCATAAAAAATCATTGCTATTGAAGGACCAGATTTAGT  
CAATGGATCCCATCAGAAAACTTTGAGGAAATCCTGCCAGAGATCATCATCTCCAAAGAGTTGAGCCTCT  
TATCCGGTGAAGTGTGCAACCGCACAGAAGGCTTTTATCGGGCTGAACCTCAGTGCAGTGTGTCCAGGGATCT  
TACCTCCTTGTCTCTCAAAGCCTGGGATCACTATAATTCACCAAAGGCTCTGCAAATGACCGGCCTGA  
TTTTTTTAAAGACGACCAGCTCTTATTGTGCTGGAATTTGAGTTTGGAGGATTGACTTAGAGCAAATGC  
GAACCAAGTTGTCTTCTTGGCTACTGCAAAGAGCATTCTACACCAGCTCACAGCCTCCCTCGCAGTGGCA  
GAGGCATCACTGCGCTTTGAGCACCGAGACTTACACTGGGGGAACGTGCTCTTAAAGAAAACAGCCTCAA  
AAAACCTCCACTACACCCTCAATGGGAAGAGCAGCACTATCCCAGCTGTGGGTGCAAGTGAGCATCATTG  
ACTACACCCTGTGCGCTTGAACGGGATGGGATTGTGGTTTTCTGTGACGTTTCCATGGATGAGGACCTG  
TTTACCGGTGACGGTGACTACCAAGTTTACATCTACAGGCTCATGAAGAAGGAGAATAACAACCGCTGGGG  
TGAATATCACCTTTATAGTAATGTGCTCTGGTTACATTACCTGACAGACAAGATGCTGAAACAAATGACCT  
TCAAGACTAAATGTAACACTCTGCCATGAAGCAAATTAAGAGAAAAATCCAGGAGTTCCACAGGACAATG  
CTGAACCTCAGCTCTGCCACTGACTTGCTCTGCCAGCACAGTCTGTTTAAGTAA

>SGK093\_WNK3\_ID#NA\_31

ATGTCCAGACTGAGGCCGACCTGGCCCTGCGGCCCCCGCCTCCTTGGCACCGCGGGGCAGCCCCGCT  
CGGGCCCCCTCCTCGCGAGCGCGCGCTTCTCCGGGAAGGCTGAGCCCCGGCGCGCTCTTCTCGTCTCA  
GCCGCGTAGCTCAGTCGACTTGGGGCTGTGAGCTTGTGCTCCCTGCCAGCCTCACCCGCTCCGGACCCC  
CCCGATCCTCCGGACTCCGCTGGTCTCGGCCCGGAGGAGCCACCGCTAGCTCAAAGAACCCCCGA  
GGGCACGTGGACCGAGGGAGCCCCGTGTGAAGGCTGCGGAAGACTCCGCGCGTCCCGAGCTCCCGGACTCTG  
CAGTGGGCCCCGGGTCCAGGGAGCCGCTAAGGGTCCCTGAAGCTGTGCCCTAGAGCGGCGGCGGAGCAG

Figure 1P

GAAGAAAAGGAGGACATGGAGACCCAGGCTGTGGCAACGTCCCCGATGGCCGATACCTCAAGTTTGACAT  
CGAGATTGGACGTGGCTCCTTCAAGACGGTGTATCGAGGGCTAGACACCGACACCACAGTGGAGGTGGCCT  
GGTGTGAGCTGCAGACTCGGAACTGTCTAGAGCTGAGCGGCAGCGCTTCTCAGAGGAGGTGGAGATGCTC  
AAGGGGCTGCAGACCCCAACATCGTCCGCTTCTATGATTCTGTGAAGTCGGTGCTGAGGGGGCCAGGTTTG  
CATCGTGCTGGTCACCGAATCATGACCTCGGGCAGCTCAAGACGTACCTGAGGCGGTTCCGGGAGATGA  
AGCCGCGGGTCCCTCAGCGCTGGAGCCGCAAAATCCTGCGGGGACTTCATTTCTACACTCCCGGGTTCCT  
CCCATCCTGCACCGGGATCTCAAGTGCACAAATGTCTTTATCACGGGACCTACTGGCTCTGTCAAAATCGG  
GGACCTGGGCCTGGCCACGCTCAAGCGCGCTCCTTTGCCAAGAGTGTATCGGGACCCCGGAATTCATGG  
CCCCGAGATGTACGAGGAAAAGTACGATGAGGCCGTGGACGTGTACGCGTTCGGCATGTGCATGTCTGGAG  
ATGGCCACCTCTGAGTACCCGTACTCCGAGTGCCAGAATGCCGCGCAAATCTACCGCAAGGTCACTTCGGG  
CAGAAAGCCGAACAGCTTCCACAAGGTGAAGATACCCGAGGTGAAGGAGATCATTGAAGGCTGCATCCGCA  
CGGATAAGAACGAGAGGTTTACCATCCAGGACCTCCTGGCCACGCTTCTTCCGCGAGGAGCGCGGTGTG  
CACGTGGAAC TAGCGGAGGAGGACGACGGCGAGAAGCCGGGCTCAAGCTCTGGCTGCGCATGGAGGACGC  
GCGGCGCGGGGGCGCCACGGGACAACAGGCCATCGAGTTCCTGTTCCAGCTGGGCCGGGACGCGGCCG  
AGGAGGTGGCACAGGAGATGGTGGCTCTGGGCTTGGTCTGTGAAGCCGATTACCAGCCAGTGGCCCGTGCA  
GTACGTGAACGGGTTGCTGCCATCCAGCGAAAGCGTGAGAAGCTGCGTAAAGCAAGGGAATTGGAGGCAT  
CCCACCAGAGCCAGGACCTCCACCAGCAACTGTGCCCCATGGCCCCCGGTCCCCCAGTGTCTTCCCCCTG  
AGCCTGAGGAGCCAGAGGCAGACCAGCACCCCTTCTCTTTCCGCCACGCCAGCTACTCATCTACCACT  
TCGGATTGCGAGACTGATGGCTACCTCAGCTCCTCCGGCTTCTTGGATGCTCAGACCCTGCCCTTCAGCC  
CCCTGGGGGGGTGCCATCCAGCCTGGCTGAGTCCCATCTCTGCCTGCCCTCGGCTTTTGCCCTATCCATT  
CACGTTCTGGCCCTGGAAGTGACTTTTCCCCGGGGACAGCTATGCCTCAGATGCAGCTTCAGGCCTTAGC  
GATGTGGGAGAAGGGATGGGACAAATGAGGAGACCCCCAGGGAGGAATCTCCGGCGCAGACCCCGATCCCG  
GCTGCGGGTCACTAGTGTCTCAGACCAGAATGACAGAGTGGTTGAGTGCCAGCTACAGACCCATAACAGCA  
AGATGGTGACCTTCCGATTTGATCTGGATGGGACAGCCCGAAGAGATTGCAGCTGCCATGGTATATAAC  
GAGTTCACTTCTGCCTTCGGAGCGAGATGGATTCTCAGACGGATTCCGGGAGATTATCCAGCGAGTGGAGAC  
CCTGTTGAAGAGAGACTGGCCCCATGGAGGCTGCTGAAGACACCCTAAGCCCCAGGAGGAGCCAGCAC  
CATTACCTGCCCTGCCCGTCCCCCTCCCAGACCCATCCAATGAAGAGCTCCAGAGCAGCACCTCCCTGGAG  
CACAGGAGCTGGACAGCCTTCTCCACCTCCTCATCTTCTCCTGGAACCTCTTGTCTCCTGGAAACCCATT  
TTCCCCTGGAACCCCCATTTCCCAGGTCCCATCTTCCCCATCACTTCTCCCCCATGTCATCCCAGCCCCCT  
CCCCATCTCCCCCATTTCTTCCCAGGTCTCCTCAAATCCCTCTCCACACCCACCAGCTCTCCACTTCCA  
TTCTCCTCCAGCACACCCGAGTTTCCGGTCCCCTCTCTCAGTGTCCCTGGAGTTCTCTCCCCACGACTTC  
TCCACTACGTTCTCTCCCACTTGTCTCAGGTCACTCTTAGTTCCCTTTCTTCTCCTCCGTGCCCCCTCA  
CTTCTTCTTCCCCCTCCACCACAGCAGCCCCCTCTCCTTTCTCTGGCTAGTGCTTCTCACTGGCTGTGATG  
ACTGTGGCCAGTCCCTGCTGTCCCCCTCAGCTGGGCTCCTTTCCAGTCTCTCCAGCCCCCTCTAGTCC  
CCTCCCTAGCCTGCCCTTCCCCCTCCCCTTGTCTCCTGGTGGCCAGGAAAGCCCTTCACCCCACACAGCTG  
AGGTGGAGAGTGAGGCCTCACACCTCCTGCTCGGCCCCCTCCAGGGGAAGCCAGGCTGGCGCCCATCTCT  
GAAGAGGGAAAGCCGAGCTTGTGGGCGTTTCCAAGTGACTTCATCCAAGGAACCGGCTGAGCCTCTTCC  
CTTGACGCCAACATCCCCCACTCTCTCTGGTTCTCCAAAACCTTCAACCCCTCAGCTCACTTCAGAGAGCT  
CAGATACAGAGGACAGTGCTGGAGGCGGGCCAGAGACCAGGGAAGCTCTGGCTGAGAGCGACCGTGACGCT  
GAGGGTCTGGGGGCTGGAGTTGAGGAGGAAGGAGATGATGGGAAGGAACCCCAAGTTGGGGGAGCCCCCA  
ACCCCTGAGCCATCCCAGCCCAGTGTGGATGAACTACTCCTACAGCAGCCTGTGTTGAGCAGCGAGGAGT  
CAGAAAGCAGTGGGGAAGATGAGGAGTTCTGGGCTGAGCTGCAGAGTCTTCGGCAGAAGCACTTGTTCAGAG  
GTGGAAACACTACAGACACTACAGAAAAAGAAATTGAAGATTTGTACAGCCGCTGGGGGAAGCAGCCCCC  
ACCGGGTATTGTGGCCCCAGCTGCTATGCTGTCCAGCCGCCAGCGCCGCTCTCCAAGGGCAGCTTCCCCA  
CCTCCCGCCGCAACAGCCTACAGCGCTCTGAGCCCCCAGGCCCTGGCATCATGCGAAGGAACTCTCTGAGT  
GGCAGCAGCACCGGCTCCCAGGAGCAGCGGCAAGCAAGGGGTGACATTCGCGGGGATGTTGGCAGGAT  
GTTTGGAGTTGTGCGCCACAGAGACGATTGAAGACGCCCTGCTTCACTTGGCCACGAGAATGAGCAAGCAG  
TGAGGGAGGCTTCGGGGCGGCTGGGCCGCTTTCAGGGAGCCCCAGATCGTGGAGTTTGTGTTTCTCCTGTCT  
GAACAATGGTGTCTGGAGAAATCTGTGAGCTACCAGGCTGTAGAAATCCTAGAAAGGTTTATGGTAAACA  
GGCAGAGAACATCTGCAGGCAAGCCACAATCCAGCCAAAGAGATAATAAGAGAGAGTCTCAGAATTGGAGGG  
CTCTGAAACAGCAGCTTGTCAACAAGTTTACTCTCCGCTTGTGTGTCATGTGTTTCCAGTGGCCAGCAAACTT  
TCCTTCCGAAACAAAATAATCAGCAACATTCCAGTCTTGAATTTCTCCAGGCTCTAGGCTATCTACACAC  
TAAAGAAGAAGTGTGGAATCAGAGCTTGATGTTTTGAAGTCCTTGAACCTCCGAATTAATCTGCCCACTC

Figure 1Q

CCCTGGCATATGTGGAGACGCTCCTAGAGGTTTTAGGATACAATGGCTGTTTGGTTCCAGCCATGAGGCTG  
CATGCAACCTGCCTGACACTGCTCGACCTGGTCTATCTTCTGCATGAACCCATATATGAGAGCCTGTTGAG  
GGCTTCAATTGAGAACTCCACTCCCAGTCAGCTGCAAGGGGAAAAGTTTACTTCAGTGAAGGAAGACTTCA  
TGCTGTTGGCAGTAGGAATCATTGCAGCAAGTGCTTTCATCCAAAACCATGAGTGTGGAGCCAGGTATGC  
ACCACTGAGCAGGACCAGCATGAGAGAGTTAAGGTGGACGAGCATGCCCTTTTGTTCATATCAGCCTGA

>SGK074\_ID#NA\_32

TCGGAGACTGACTGCTACGACATCATCGAGGTCTTGGGCAAGGGGACCTTCGGGGAGGTAGCCAAGGGCTG  
GCGGCGGAGCACGGGCGAGATGGTGGCCATCAAGATCCTCAAGAATGACGCCTACCGCAACCGCATCATCA  
AGAACGAGCTGAAGCTGCTGCACTGCATGCGAGGCCTAGACCCTGAAGAGGGCCACGTTCATCCGCTTCCTT  
GAGTTCCTTCCATGACGCCCTCAAGTCTACCTGGTCTTTGAGCTGCTGGAGCAAAACCTTTTCGAGTTCCA  
GAAGGAGAACAACCTTCGCGCCCCCTCCCCGCCGCCACATCCGTACAGTCACCCTGCAGGTGCTCACAGCCC  
TGGCCCCGGCTCAAGGAGCTGGCTATCATCCACGCTGATCTCAAGCCTGAGAACATCATGCTGGTGGACCAG  
ACCCGCTGCCCCCTTCAGGGTCAAGGTGATTGACTTCGGATCCGCCAGCATTTCAGCGAGGTGCGCTACGT  
GAAGGAGCCATACATCCAGTCGCGCTTCTACCGGGCCCCCTGAGATCCTGCTGGGGCTGCCCTTCTGCGAGA  
AGGTGGACGTGTGGTCCCTGGGCTGCGTCATGGCTGAGCTGCACCTGGGCTGGCCTCTCTACCCCGGCAAC  
AACGAGTACGACCAGGTGCGCTACATCTGCGAAACCCAGGGCCTGCCCAAGCCACACCTGTTGCACGCCGC  
CTGCAAGGCCCAACCACTTCTTCAAGCGCAACCCCCACCTGACGCTGCCAACCCCTGGCAGCTCAAGTCTCT  
CGGTGACTACCTGGCCGAGACGAAGGTGCGCCCATTTGGAGCGCCGCAAGTATATGCTCAAGTCTGTTGGAC  
CAGATTGAGACAGTGAATGGTGGCAGTGTGGCCAGTCGGCTAACCTTCCCTGACCGGGAGGCGCTGGCGGA  
GCACGCCGACCTCAAGAGCATGGTGGAGCTGATCAAGCGCATGCTGACCTGGGAGTCAACGAACGCATCA  
GCCCCAGTGCTGCCCTGCGCCACCCCTTCGTGTCCATGCAGCAGCTGCGCAGTGCCACGAGACCACCCAC

>SGK087\_ID#NA\_33

ATGAACACCCCTCTGCAGTCTGCACATACACACCTTACCCTGGTTCGAGTTCATAGCCTACTTGGTGTCTGA  
CTCCCCCTTTCTGCTGCTTCAACCTTGTCTCTGCCCACGTGCAACTTTGATCCCTGCCCCACTAAACTGG  
GTGCTGGCACAATGCCTCCCGCAGCGGCGGGCGGTGAGCGCCGGCCTCATGCAGCTCCTCCCGCCGCTAT  
CCGCACCGGAACAAAGTAAGGGCCGCGAGGCTCGTACTTCACGAGCAGTCAGGCGCGAGTACGAGGCAGG  
CGACGAAGCTTCTGGTCGACGGGGTCTGGGGGAGAAATGAAAAGCGTGCAGAAATGCAAGAGCTGACCGGG  
GAATGTGGGTGAGAGAGAGGCGGGCTGTGGTGC CGGACAAGAAAATACAGTTGTAGTTTTTGCAGCAGCT  
GCATCCCCGGAATCCGACTTCTCCAAATCGTCTTCGTGATATTGCCAGAGAAATGGTACCAGAATACAA  
ATCTGATCTTCAGCCCAGGACTCAAATGGATGCTAAAAAGCCAAGGAAATGTGATTTGACTCCCTTCTCTGG  
TTTTGAAAGCAAGAAAGAAACAAAAGTTACCTCTGCGAAGCACCTCTGGGTACAGCGCAAAGGCAAAGGT  
GTGGAGGATCATGCTGATTGGCAGGGCCAAATCTGGGTGGGTGTGGAGGGAAAATCAGCAGAAGGGCAAAGG  
CCTTAATCTAACAGGTGCATTTAACAGAACTCCAGAAAACAAAGTATCTCCCCAAGGCCCTGCACTGAATG  
CATTAGAAATTGACTCCAGCTATCTGGACCCCCAGGCTCTTGCCCTTCTCAAGGATGGCAGGGCCTCCTC  
TTTCCACCTCTACTAGGGACAGAGCTGGTTGGAAGTAAACTTTCGGTTCAGATCCAGAAGCCACCTTCCAA  
TATCAAGAACTCCAGAATGACCCAAGTCTTTCATAAGAACACCAAGTGTACTTCACTCCCCTTTGTGGACA  
CCAAGGGGAAGAAGAAATACGGTAAGCTTCCCACACATTAGCAAGAAAGTCTGCTGAAGTCAATCCCTGCTG  
TATCAGGTGAGTGCAGACGACTGGACCTGAGAAGGCAGGTGAAAGACCTGGTTTTAGTCCAAAGCCAGT  
TTTAGAAAGGCCCAGGATAGTGGGGAAAAGCACAGTAGCTGCTGAAGAGGAGAATCAAGCTCACAATCAGA  
TGCCGGCCTCAGAGCTCAAGGCTTCAGAAATACCTTTCCACCCTAGCATTAAAACCCAGGATCCCAAGGCA  
GAGGAGAAGTCAACAAAGAAGCAAAAGGTGACTCTGACAGCGGCAGAGGCCCTAAAGCTTTTTAAGAACCA  
GCTGTCTCCATATGAACAAAGTGAAATCCTGGGCTACGCGGAGCTGTGGTTCTGGGTCTTGAAGCCAAGA  
AGCTCGACACGGCTCCTGAGAAATTTAGCAAGACGAGTTTTGATGATGAGCATGGCTTCTATCTGAAGGTC  
CTGCATGATCACATTGCCTACCGCTATGAAGTTCTGGAGACAATCGGGAAGGGTCTTTTGGACAGGTGGC  
CAAGTGCTTGGATCACAAAAACAATGAGCTGGTGGCCCTGAAAATCATCAGGAACAAGAAGAGGTTTCAAC  
AGCAGGCCCTGATGGAGCTGAAGATCCTGGAAGCTCTCAGAAAGAGGACAAAGACAACACCTACAATGTG  
GTGATGAGTTGATGAAGAATAACAATTTCAAGGCTTCAGTCTGTCCATAGTTCGGCGCTTCACTCTCTCTG  
TTTTGAAGTGCTTGCAGATGCTTTCGGTAGAGAAAATCATTCACTGTGATCTCAAGCCGAAAATATAGTG  
CTATACCAAAAGGGCCAAGCCTCTGTAAAGTCAATTGACTTTGGATCAAGCTGTTATGAACACCAGAAAGG

Figure 1R





>SGK125\_MY03A\_ID#NA\_36  
ATGTTTCCATTAATTGGAAAAACAATCATCTTTGATAACTTTCTCTGATCCTTCTGATACATGGGAAATCAC  
TGAGACAATTGGCAAAGGAACCTTATGGGAAAGTTTTTAAAGTATTGAATAAGAAAAATGGCCAAAAAGCAG  
CAGTCAAAATTCTTGATCCAATTCACGATATTGACGAAGAGATTGAAGCAGAATATAACATCTTAAAAGCA  
CTTTCTGACCACCCTAATGTGGTCAGATTCTATGGGATATACTTTAAGAAGGATAAAGTAAATGGAGACAA  
GCTGTGGTTGGTTCTTGAGCTCTGCAGTGGAGGATCAGTGACTGACCTTGTGAAAGGATTTCTGAAGAGGG  
GTGAAAGAATGAGTGAGCCTCTAATTGCCTATATTTTACATGAAGCACTAATGGGACTTCAACATTTGCAT  
AACAACAAACTATCCACAGAGATGTGAAAGGCAATAACATTCTATTGACCACGGAAGGTGGAGTGAAACT  
AGTAGATTTTGGTGTGTCTGCACAGCTCACCAGTACCCGGCACCGTCGGAACACATCCGTAGGAACACCGT  
TTTGGATGGCTCCTGAGGTGATTGCATGTGAACAGCAATTGGATACCACTTATGACGCCAGATGTGACACT  
TGGTCCCTGGGTATCACGGCCATTGAGCTGGGTGATGGAGATCCTCCACTAGCTGACCTTCATCCCATGAG  
AGCACTCTTCAAAATACCAAGGAATCCACCCCAAACTAAGGCAGCCTGAGCTATGGTCAGCAGAATTCA  
ATGACTTCATAAGCAAGTGCTTGAATAAGATTATGAAAAGCGTCCAACAGTGTGAGAATTTTACAGCAT  
AAATTCACTACTCAAATTGAGGGCAAAGATGTGATGCTACAAAACAATAACGGAATTCATTGGCATCCA  
TCAATGCATGGGAGGCACAGAAAAGGCCAGACGTGAACGTATTACACGAAGAAAGGGAACCTCAACCGAC  
CTCTAATATCCAATCTGAAGGATGTAGATGATTTAGCAACCCTAGAAATTTTGGATGAGAATACAGTCTCA  
GAGCAACTTGAGAAGTGTTATTCCAGAGATCAGATCTACGTCTATGTGGGAGACATACTCATTGCTCTTAA  
CCCTTTTCAGAGTCTGGGTCTTTACTCCACAAAGCATTCCAACTATATATTGGATCAAAGAGAAGTGCCA  
GTCCTCCTCACATTTTTCGAATGGCTGACTTAGGATATCAATCTATAATAACATATAATTGAGATCAGTGC  
ATTGTTATTTCTGGAGAAAGTGGTGCTGGAAAGATGAAAATGCTCATCTTTTAGTTTCAGCAGCTGACAGT  
GCTTGGAAAAGGCTAATAACAGAACCTTGAAGAGAAGATTTTACAAGTGAACAATTTGGTAGAAGCCTTTG  
GCAATGCCTGCACTATTATAAATGACAATTCTAGCAGATTTTGGAAAATACTTAGAAATGAAATTCACCTCT  
TCTGGAGCGGTAGTGGGAGCACAGATTTCTGAATATCTCCTGGAATAATCCCGAGTTATCCACCAAGCTAT  
TGGAGAAAAAATTTTCATATTTTACTACATTTATGCTGGTTTGGCTGAAAAGAAGAACTAGCCCCATT  
ACAACTGCCTGAAAATAAGCCTCCCAGGTACCTACAAAATGACCACCTCAGAACAGTACAAGACATCATG  
AATAATAGTTTCTATAAATCCAGTATGAATTAATTGAGCAATGTTTCAAAGTCATAGGTTTACAATGGA  
GCACTTGGTAGTATATACAGCATACTCGCTGCAATCTTGAATGTTGGCAACATTGAATTTTCTTCTGTGG  
CACTGAACACCAGATTGACAAGAGCCACATTTCTAATCATACAGCCCTGGAGAAGTGTGTTTCTTGTGCTT  
TGCATTGGGCAGATGAGCTACAGAAGCTCTCACCTCCACTGTGTGGTCACTAGAGGAGAAAACAATTAT  
ACGACCCAATACTGTAGAAAAAGCTACCGATGTGAGGATGCCATGGCTAAAACCTTATATGGACGTCTCT  
TTAGTTGGATAGTCAATTGCATTAACAGTTTGTGTAAGCATGACTCATCACCAGTGGGAATGGTGATGAG  
CTGAGCATTGGCATTCTTGATATATTTGGTTTTGAAAATTTCAAAAAAATTCCTTCGAGCAGCTGTGCAT  
TAACATTGCAAATGAACAAATTCAGTATTATTATAATCAACATGTGTTTGCATGGGAACAGAATGAATACC  
TAAATGAAGATGTGGATGCTAGAGTTATTGAATATGAGGATAAAGTGGCCCTCTTAGATATGTTTCTGCAA  
AAGCCAATGGGTTTACTTTCCCTACTTGATGAAGAAAGTAGATTTCCCAAGGCCACTGACCAGACTCTTGT  
AGAAAAATTTGAAGGTAACCTGAAATCACAATACTCTGGAGACCCAAAAGAATGGAACCTTAGTTTTGGAA  
TTCACCATTATGCAGGAAAGGTCTCTATAATGCAAGTGGATTCTTAGCCAAAAACAAGACACTAGTTCTCT  
ACTGACATTGTGCTACTTTTGGAGTCACTCCGACAAAGTGAATTTAGGCAACTAGTCAACCACCTCTGAC  
CAAAACAGGTAATCTGCCACATTCTAACTAAAAATGTTAAACTATCAAATGAGGACTTCAGAAAAATTAA  
TCAACCTGGCAAAGGGCGACACTGGAGAAGCCACAGTCATGCCAGAGAGACAACCAACATGAAAACACAA  
ACGTTTGCATCATATTTTAGATATTCCCTGATGGATTGTTGTCTAAAATGGTGGTGGGCCAACCTCATTT  
TGTCGGTTGCATCAAACCAATAGTGAGCGTCAGGCAAGAAAATATGACAAAAGAGAAAGTTCTGCTACAGC  
TTCCGTACACAGGAATCTGGAAACAGCAAGAATTCGAAGACTAGGATTCTCCCATCGGATACCTTTTGTCT  
AACTTTATAAAGCGGTACTACCTTCTCTGCTACAAGTCGAGCGAGGAGCCCCGATGAGCCCTGACACCTG  
TGCCACCATTTTGGAAAAAGCTGGTCTCGATAACTGGGCTCTTGGAAAAACAAAAGTGTTCCTTAAGTATT  
ATCACGTGGAGCAGTTAAATCTAATGCGAAAGGAAGCTATTGACAAGCTTATTTTGATTCAAGCTTGTGTC  
AGAGCATTTCTTGTGTTCAAGAAGATACCAAAAAATACAGGAGAAAAGGAAAGAAAGCGCTATAATAATACA  
GTGAGCTGCAAGAGGACACCTTGTGCGGAAACAAAGAAAAGAAATTGTTGACATGAAAAACACAGCAGTAA  
CAACATTCAAACCTCTGATCAGGAATTCGACTACAAGAAAACCTTTGAAAAATACAAGGGAATCTTTCGTG  
AAGAAACAAGCAGAAAAATGCAATCTCTGCTAATGAAAGATTCAATTCAGCTCCAAATAATAAGGAAGTGT  
ATCTGTAGTGAAGACTTCCACTTTCAAACCTGAAGAGGAAACCACCAATGCTGTGGAGAGTAACAACAGAG  
TGTATCAGACTCCAAAAAATGAATAATGTGTATGAGGAAGAGGTTAAGCAAGAATTTTACCTTGTAGGG  
CCAGAAGTAAGCCCCAACAGAAGTTTGTCAAAGACCTGGAAGAGAACAGCAATCTAAGGAAAGTGGAGAA

Figure 1T



AGAGGAAGCTATGATCCAGAGTTACTATCAGAGGTACACAGAGGAGAGGAATTGTGAAGAGTCAAAAGCAG  
CATATCTAGAAAGGAAGGCCATATCAGAAAGGCCAAGCTACCCAGTGCCTTGGTTAGCTGAAAATGAGACT  
TCCTTTAAAAAACTTTGGAACCTACACTTAGCCAAAGGTCAATTTATCAAAATGCAAACAGCATGGAAAA  
AGAAAAGAAGACATCTGTAGTTACCCAGAGTGCACCGATATGCAGCCAGGAGGAAGGCAGAGGCCGTCTGA  
GGCATGAGACAGTCAAAGAGAGGCAAGTTGAACAGTGACACAGGCCCAGGAGGAAGAAGATAAAGCAGCG  
GTATTCATTTCAGAGCAAATACCGGGGTTACAAGAGAAGGCAGCAGTTGAGGAAGGACAAGATGTCTTCTTT  
TAAGCATCAGAAGATTGTCAACACCAACAGAAGTAGCAAGAAACACTCATAATTTGTATTCTCTATCCCA  
CAAAACATGAGGAAATCAATAACATCAAGAAGAAGGATAACAAAGACTCGAAAGCAACTTCAGAAAGAGAA  
GCATGTGGTTTGGCAATTTTTCAAAACAGATATCAAAGTTATCTGAAGAATATTTCAATTCAGCAAAAA  
ATTGAATGAAATGATTTTGTACAGCAACTGAAGTCACTTTATCTGGGTGTCTCGCACCATAAGCCAATTA  
ATAGACGAGTTTCTTCTCAGCAGTGCCTCTCAGGTGTCTGTAAAGGAGAGGAGCCAAAAATATTGAGACCC  
CCAAGACGACCCCGGAAACCCAAAACATTAAATAACCTGAAGACTCCACATACTATTATCTACTTCATAA  
GTCAATCCAAGAAGAAAAACGAAGACCAAGGAAGACAGTCAGGGAAAATTATTAGATTTGGAAGATTTCT  
ATTATAAGGAATTTTGGCCAGTCGTTCTGGACCAAGGAACATAGCCCTAGTTTAAGAGAACGAAGACCA  
CAGCAAGAACTCCAGAATCAATGTATTAAAGGCTAATGAAAGGTGCTGGGCGGCGGAGAGCCCCGAGAAGGA  
GGAGGAGAGAGAGCCAGCAGCAACCCCTACGACTTCAGGAGGCTCCTGCGCAAAACCTCCCAGCGCCGGC  
GCCTCGTCCAGCAGTCCTAA

>SGK445\_ID#NA\_37

TTTAAAGTTGGAAATCTGCCTGGTAAGGGATCATTGCTGGTGTCTGCAAAGTTGAGTCCATACACACTGG  
TTTGGAAATTTTCAGTCCAGATGATAGTTAAGAAAGCAGTAAGAATACAGAGAGTCCACAATGAGATGAAAA  
TGCCTGCCAGTTGAAACATCCTCCTACACTGGAGCTTTATAAATATTTTAAAGACAAGGAT

>SGK127\_ID#NA\_38

ATGACGAAAAGCGAGGAGCAGCAGCCTCTGAGTTTGCAAAAAGCCTTACAGCAGTGCGAACTGGTCCAAAA  
CATGATAGACTTGAGCATCTCCAACCTGGAAGGGCTTAGGACCAAATGTGCTACCTCCAACGACCTCACAC  
AAAAAGAAATCCGGACCTTGAGAGCAAGCTGGTGAAGTACTTCAGCCGGCAGCTGTCTTGCAAAAAGAAG  
GTAGCCTTGCGAGGAGCGCAACGCGGAGCTGGACGGCTTCCCCCAGCTACGGCAGTGGTTCCGAATCGTCGA  
TGTGCGCAAGGAGGTCTTGAGGAAATCTCCCCGGCCAGCTGAGCCTGGAGGACCTCTTGAGATGACGG  
ATGAACAGGTGTGCGAGACTGTGGAGAAATACGGAGCCAACCGGGAGGAGTGTGCCCGCTCAACGCCTCC  
CTCTCCTGCCCTCAGGAATGTCCACATGTTCAGGAGGCAACCTTTCCAAACAAGACTGGACCATCCAGTGGCC  
CACCACAGAGACGGGGAAGGAGAACAATCCCGTGTGCCCCCGGAGCCACCCCGTGGATCCGCAACCCATC  
TCTCCAGAGCCCCAGGGTCCCGTCCAAGTGCGTCCAGCACTATTGTACACACCCCACTCCCGGGGCC  
CCTGTGTACACCCACGTGGACAGGCTTACCGTGGACGCTACCCGGGCTTGTGCCCGCCCCCGCACTGGA  
GTGGGGCCACCGTTCCCTGCCCCCATCGCCCCGCGCAGCGGCACGCGGTCCGCACCCCGCGCGCACCCCCA  
ACATCGTCAACACCGTGACCCCGCGGGCACGCGCCCATGAGGAAGAAGAACAAGCTGAAGCCCCCGGGG  
ACCCACCGCCCTCCTCCGAAAACCTGATACACTTGATCCCGGATTACCGCGCTGCATCGGAGCAAATC  
CCACGAGTTCCAGCTGGGGCACCGCGTGGACGAGGCCACACGCCCAAAGCCAAGAAGAAGAGCAAACCT  
TGAACCTCAAGATCCACAGCAGCGTAGGCAGCTGCGAGAACATCCCTCTCAGCAGCGCTCCCCGTGCTG  
TCCGAGCGCTCCCTCCGCTCCTTCTTTGTGGGACACGCACTTTCTGCTTCCACCCCTCCTGTTACAC  
TGAGGCCAATTTCTGCAACACACTGTCACTGCCACGCTGGTCCCCGAGATCCCTCGCAGAGATCTTG  
GCAACTCCATCAAGCACAGGTTTTCACCAAGTACTGGATGTCTCAGACGTGCACAGTCTGTGGGAAAGGG  
ATGCTTTTGGCCTCAAGTGTAAAACTGCAAGTTAAAGTGCCACAACAATGCACCAAGAAGCCCCACC  
CTGTATCTTCTGATCATCCACCGAGGAGATCCAGCAAGGTTAGTCCGGACAGAGTCCGTTCCGTGTGACA  
TCAACAACCTCTACGGAAGCCACCTCGCTATTTCAGACCTGCACATCAGTCAGACGCTCCCCAAAACCAAC  
AAAATCAACAAGGACCACATCCCTGTCCCTTACCAGCCAGACTCCAGCAGCAACCCCTCCTCCACGACGTC  
CTCCACGCCCCCTCGCCAGCACCCCCCTCCCTCCTAGTGCCACGCGCCCTTCTCCCTACACCTTCCC  
CACAGTGCACACGGCAGCAGAAGAACTCAACCTGCCAGCATCCCACTACTACAAATACAAGCAGCAGTTT  
ATCTTCCAGATGTGGTGCCGGTGCCGAGACGCCGAGCCCGGGCGCCAGGTCATCTGCACTCCGGTGAC  
CTCGAATCCAATCTTGGAAAGGAAATCCATTACTTCAAAATGAAGTGAGGCCAACGTCGGAGAATGAAGAGG  
TCCATGATGAGGCCGAAGAGTCAGAGGATGACTTCGAGGAGATGAACCTGTCCCTCTCTCGGCCCCGAGC  
TTCCACGCAAGGCCAGCCAGACCAGCATCTTCTTCAGGAGTGGGACATCCCTTTGAGCAGCTGGAGAT  
CGGCGAGCTCATTGAAAGGGCCGCTTTGGGCAAGTGTAACACGGCCGCTGGCATGGCGAGGTGGCCATCC

Figure 1U

GGCTGATTGACATTGAGAGGGACAACGAGGACCAGCTCAAGGCCCTTCAAGCGGGAGGTGATGGCCTACAGG  
CAGACACGGCATGAGAACGTGGTGTCTTTTCATGGGTGCCTGCATGAGCCCGCTCACCTGGCCATCATCAC  
CAGCCTCTGTAAGGGACGGACGCTCTATTCCGTTGTGAGGGATGCCAAAATCGTTTTGGATGTCAACAAAA  
CCAGGCAGATTGCTCAAGAAATTGTGAAGGGCATGGGCTACCTCCACGCCAAGGGAATCCTACACAAGGAC  
CTCAAGTCAAAGAACGTCTTCTATGACAACGGCAAAGTGGTCATCACGACTTTGGACTCTTCAGCATTTC  
TGGGGTGCTGCAGGCTGGCAGGCGGAGGACAAACTGCGCATCCAGAATGGCTGGCTATGCCACCTGGCAC  
CAGAGATCATCCGCCAGCTGTCCCCCGACACAGAGGAGGATAAGCTCCCCCTTCTCCAAGCACTCTGACGTC  
TTTGGCCCTTGGCACAATCTGGTATGAACTCCACGCCAGGGAATGGCCTTTCAAGACCCAACCAGCAGAGGC  
AATAATCTGGCAAATGGGCACAGGCATGAAACCCAACCTCAGCCAGATTGGCATGGGAAAAGAAATCTCGG  
ACATTCTTCTTCTGCTGGGCCCTTTGAACAAGAAGAGAGACCTACCTTCACCAAGCTCATGGACATGCTG  
GAGAAACTGCCAAAGCGAAACCGTCGCCTGTCTACCCTGGACATTTCTGGAAGTCTGCAGAGCTGTGA

>SGK009\_ANKRD3\_ID#NA\_39

ATGAGGGCGACGGCGGGACCCATGGGCCCTGGCGCTGCTGCGCACCTTCGACGCGGGCGAGTTCACGGG  
CTGGGAGAAGGTGGGCTCGGGCGGCTTCGGGCAGGTGTACAAGGTGCGCCATGTCCACTGGAAGACCTGGC  
TGCCCATCAAGTGCTCGCCAGCCTGCACGTGACGACAGGAGCGCATGGAGCTTTTGGAGAAGCCAAG  
AAGATGGAGATGGCCAAGTTTCGCTACATCCTGCCTGTGTATGGCATCTGCCGCGAACCTGTGCGCCTGGT  
CATGGAGTACATGGAGACGGGCTCCCTGGAAAAGCTGTGCGCTCGGAGCCATTGCCATGGGATCTCCGGT  
TCCGCATCATCCACGAGACGGCGGTGGGCATGAACTTCTTGCATGTCATGGCCCCGCCACTCTCGCTG  
GACCTCAAGCCCGCAACATCCTGCTGGATGCCCACTACCACGTCAAGATTTCTGATTTTGGTCTGGCCAA  
GTGCAACGGGCTGTCCCACTCGCATGACCTCAGCATGGATGGCCTGTTTGGCACAATCGCCTACCTCCCTC  
CAGAGCGCATCAGGGAGAAGAGCCGGCTCTTCGACACCAAGCACGATGTATACAGCTTTGCGATCGTCATC  
TGGGGCGTGCTCACACAGAAGAAGCCGTTTGCAGATGAGAAGAATCCTGCACATCATGGTGAAGGTGGT  
GAAGGGCCACCGCCCCGAGCTGCCGCCCGTGTGCAGAGCCCGCGCGCCTGCAGCCACCTGATACGCC  
TCATGCAGCGGTGCTGGCAGGGGGATCCGCGAGTTAGGCCACCTTCCAAGGAAACGGGTGAATGGGGAG  
CTCATCCGCCAGGTGCTGGCAGCTCTGCTCCCTGTGACTGGCAGGTGGCGCTCCCCGGGGAAGGCTTCCG  
CCTTGAGTCTGAAGTCATCATCCGAGTGACATGTCCCTTTCTTCCCCACAAGAAATTACTTCTGAAACCG  
AGGACCTGTGTGAAAAGCCTGATGACGAAGTGAAGAAACTGCTCATGATCTGGACGTGAAAAGCCCCCG  
GAGCCCAGGAGCGAGGTGGTGCCTGCGAGGCTCAAGCGGGCCTCTGCCCCACCTTCGATAACGACTACAG  
CCTCTCCGAGCTTCTCTCACAGCTGGACTCTGGAGTTTCCAGGCTGTGAGGGCCCCGAGGAGCTCAGCC  
GCAGTCTCTCTTCCAGAGGATCACTGTGCTGTCTTGTGAGCGGGAACCTTCAACCAGCGATCTGGGCAC  
CACAGACGTCCAGAAGAAGAAGCTTGTGGATGCCATCGTGTCCGGGGACACCAGCAAATGATGAAGATCC  
TGCAGCCGCAGGACGTGGACCTGGCACTGGACAGCGGTGCCAGCCTGCTGCACCTGGCGGTGGAGGCCGG  
CAAGAGGAGTGCGCCAAGTGCTGCTGCTCAACAATGCCAACCCCAACCTGAGCAACCGTAGGGGCTCCAC  
CCCGTTGCACATGGCCGTGGAGAGGAGGGTGGCGGGTGTGCTGGAGCTCCTGTGGCACGGAAGATCAGTG  
TCAACGCCAAGGATGAGGACCAGTGGACAGCCCTCCACTTTGCAGCCAGAACGGGGATGAGTCTAGCACA  
CGGCTGTGTTGGAGAAGAACGCTCGGTCAACGAGGTGGACTTTGAGGGCCGGACGCCCATGCACGTGGC  
CTGCCAGCACGGGCAGGAGAATATCGTGCATCCTGCTGCGCCGAGGCGTGGACGTGAGCCTGCAGGGCA  
AGGATGCCTGGCTGCCACTGCACTACGCTGCCTGGCAGGGCCACCTGCCCATCGTCAAGCTGCTGGCCAAG  
CAGCCGGGGGTGAGTGTGAACGCCCAGACGCTGGATGGGAGGACGCCATTGCACCTGGCCGCACAGCGCGG  
GCACTACCGCGTGGCCCGCATCCTCATCGACCTGTGCTCCGACGTCAACGTCTGCAGCCTGCTGGCACAGA  
CACCCCTGCACGTGGCCCGGAGACGGGGCACACGAGCACTGCCAGGCTGCTCCTGCATCGGGGCGCTGGC  
AAGGAGGCCGTGACCTCAGACGGCTACACCGCTCTGCACCTGGCTGCCCGCAACGGACACCTGGCCACTGT  
CAAGCTGCTTGTGAGGAGAAGGCCGATGTGCTGGCCCGGGGACCCCTGAACCAGACGGCGCTGCACCTGG  
CTGCCGCCACGGGCACTCGGAGGTGGTGGAGGAGTTGGTCAGCGCCGATGTCAATTGACCTGTTTCGACGAG  
CAGGGGCTCAGCGCGCTGCACCTGGCCGCCAGGGCCGGCACGCACAGACGGTGGAGACTCTGCTCAGGCA  
TGGGGCCACATCAACCTGCAGAGCCTCAAGTTCAGGGCGGCCATGGCCCCCGGCCACACTCCTGCGGC  
GAAGCAAGACCTAG

>SGK421\_STK22A\_TSK1\_ID#NA\_40

ATGGATGACGCTGCTGTCTCAAGCGACGAGGCTACCTCCTGGGGATAAAATTTAGGAGAGGGCTCCTATGC  
AAAAGTAAAATCTGCTTACTCTGAGCGCCTGAAGTTCAATGTGGCGATCAAGATCATCGACCGCAAGAAGG

Figure 1V

CCCCCGCAGACTTCTTGGAGAAATTCCTTCCCCGGGAAATTGAGATTCTGGCCATGTTAAACCACTGCTCC  
ATCATTAAGACCTACGAGATCTTTGAGACATCACATGGCAAGGTCTACATCGTCATGGAGCTCGCAGTCCA  
GGGCGACCTCCTCGAGTTAATCAAAACCCGGGGAGCCCTGCATGAGGACGAAGCTCGCAAGAAGTTCCACC  
AGCTTTCTTGGCCATCAAGTACTGCCACGACCTGGACGTCTGCCACCGGACCTCAAGTGTGACAACCTT  
CTCCTTGACAAGGACTTCAACATCAAGCTGTCCGACTTCAGCTTCTCCAAGCGTGCCTGCGGGATGACAG  
TGGTCGAATGGCCTTAAGCAAGACCTTCTGTGGGTACCAGCGTATGCGGCCCCAGAGGTGCTGCAGGGCA  
TTCCCTACCAGCCCCAAGGTGTACGACATCTGGAGCCTAGGCGTGATCCTCTACATCATGGTCTGCGGCTCC  
ATGCCCTACCAGCACTCCAACATCAAGAAGATGCTGCGTATCCAGAAGGAGCACCGCGTCAACTTCCCACG  
CTCCAAGCACCTGACAGGCGAGTGCAAGGACCTCATCTACCACATGCTGCAGCCCGACGTCAACCGGCGGC  
TCCACATCGACGAGATCCTCAGCCACTGCTGGATGCAGCCCCAAGGCACGGGGATCTCCCTCTGTGGCCATC  
ACAAGGAGGGGGAGAGTTCCCGGGGAACTGAACCCCTTGTGGACCCCCGAACCTGGCTCTGACAAGAAGTC  
TGCCACCAAGCTGGAGCCTGAGGGAGAGGCACAGCCCCAGGCACAGCCTGAGACAAAACCCGAGGGGACAG  
CAATGCAAATGTCCAGGCAGTCGGAGATCCTGGGTTTCCCCAGCAAGCCGTCGACTATGGAGACAGAGGAA  
GGGCCCCCCCCAACAGCCTCCAGAGACGCGGGCCAGTGA

>SGK047\_ID#NA\_41

CCCAGGGAGGCTGTCTCCATTGTGTCTCTCTGCTGGGTGAGGCCCCCGGCCACCAGACTCCGAGCACTTAA  
GATTCTGACCCATCCCGCATT

>SGK196\_ID#NA\_42

ATGGAAAAGCAGCCCCAGAACAGCAGGAGAGGCCTCGCCCCCGAGAGGTGCCGCCAGCTGTTGGGCTGCT  
GCTGATCATGGCCCTGATGAATACTCTACTCTACCTCTGCCTCGACCACTTCTTCATCGCTCCTCGACAAT  
CCACTGTGGACCCACACACTGTCCCTATGGTCACTTCAGGATAGGACAGATGAAAACTGCTCACCTTGG  
CTGTCTGCGAGGAGCTGAGAACAGAAGTGAGACAGCTGAAGCGTGTGGGGAAGGAGCTGTAAAGAGAGT  
CTTTCTGTCTGAGTGGAAGGAGCACAAAGTTGCACTCTCACAGCTCACCAGCCTGGAGATGAAAGATGATT  
TCCTCCATGGACTGCAGATGCTGAAATCTCTCCAAGGCACACATGTTGTCACGCTGCTTGGCTATTGTGAG  
GATGACAACACTATGCTTACTGAATATCACCTCTAGGTTTCTTGAGTAACCTGGAAGAAACACTAAACCT  
TTCAAAGTACCAAAATGTGAACACGTGGCAGCACAGGCTGGAGCTGGCCATGGACTATGTGAGCATCATTA  
ATTACCTGCACCACAGCCCTGTGGGCACACGGGTGATGTGCGACTCCAACGACCTGCCGAAGACACTGTCC  
CAGTATCTGCTAACAAGCAACTTCAGCATTTTGGCAAATGACTTGGACGCCCTTACCCTTGGTGAACCACAG  
CTCCGGGATGCTGGTGAAGTGCGGCCACAGGGAGCTGCATGGGGATTTCGTGGCTCCAGAGCAACTGTGGC  
CCTATGGAGAGGACGTGCCTTTCCACGATGATCTCATGCCCTCATATGATGAGAAGATTGACATTTGGAAG  
ATCCAGACATCTCCAGTTTCTTCTGGGGCACATGAAGGGAGTGATATGGTCCGATTCCATTTGTTTGA  
TATTCACAAAGCATGCAAGAGCCAGACTCCCTCAGAAAGACCCACTGCCAGGACGTTCTGGAGACCTACC  
AGAAGGTCTTGGATACACTTAGAGATGCCATGATGTCTCAGGCAAGAGAGATGCTGTGA

>SGK396\_ID#NA\_43

ATGGATAATATTGATGAAATCCTAGAGAAGACTGAGTCAAGTGTCTGCAAAGAGCTGGAGATAGCTCTGGT  
TGATCAAGGTGATGCAGACAAGGAGATAATTTCAAATACATATAGTCAAGTACTGCAAAAGATTCAATTCAG  
AGGAAAGGCTCATTGCCACAGTACAAGCTAAGTACAAGGACAGTATTGAGTTTAAAAAGCAGCTTATTGAA  
TATTTAAAGAAGATTCCCAGTGTGGATCACTTGCTATCCATTAAGAAGACATTGAAAAGCTTAAAGCTCT  
ACTCAGATGGAAATTGGTTGAAAAGAGTAATTTGGAAGAGTCAGATGATCCTGATGGCTCTCAAATTGAGA  
AAATAAAAGAAGAAATAACTCAGCTGCGCAATAATGTCTTTTCAGGAAATTTATCATGAGAGAGAGGAATAT  
GAGATGCTAACTAGTTTGGCACAGAAATGGTTCCCTGAGCTGCCTCTGCTTCATCCTGAAATAGGATTACT  
CAAATACATGAACTCTGGTGGTCTCCTTACAATGAGCTTGAACGAGATCTTCTTGATGCTGAGCCCATGA  
AGGAACTTAGCAGCAAGCGTCTCTTGGTACGTTCTGAGGTTAATGGGCAGATAATTCGTAAAGGGCTAT  
TCTGTGGATGTTGACACAGAAGCCAAGGTGATTGAGAGAGCAGCCACCTACCATAGAGCTTGGAGAGAAGC  
TGAAGGAGACTCAGGGTTACTGCCATTGATATTCCTGTTTTATGTAAGTCTGATCCTATGGCTTATCTGA  
TGGTCCCATACTACCCTAGGGCAAACCTGAATGCTGTTCAGCCAACATGCCCTTTAAATTCAGAAGAACT  
TTAAAGGTCATGAAAGGTGTTGCCAGGGTCTGCATACATTCATAAGGCTGACATAATTCATGGATCACT  
TCATCAGAACAATGATTTGCTTTAAACCGTGAACAAGGAATTGTTGGAGATTTTGACTTCACCAAATCTG  
TGAGTCAGCGAGCCTCGGTGAACATGATGGTTGGTGAAGTTGAGTTTGGATGTCACCTGAGTTGAAAATGGGA  
AAACCTGCTTCTCCAGGTTCCAGACTTATATGCTTATGGCTGCCTCTTATTATGGCTTTCTGTTCAAATCA

Figure 1W

GGAGTTTGAGATAAAATAAAGATGGAATCCCCAAAGTGGATCAGTTTTCATCTGGATGATAAAGTCAAATCCC  
TCCTCTGTAGCTTGATATGTTATAGAAGTTCAATGACTGCTGAACAAGTTTAAATGCTGAATGTTTCTTG  
ATGCCAAAGGAGCAATCAGTTTCAAACCCAGAAAAAGATACTGAATACACCCTATATAAAAAGGAAGAAGA  
AATAAAGACGGAGAACTTGGATAAATGTATGGAGAAGACAAGAAATGGTGAAGCCAACCTTTGATTGTTAA

>SGK279\_PKN\_ID#NA\_44

ATGAGCGTGGGCTGCCAGAGCCTGAGCCGCCCCGCTCCCTGACCTGCTGTGGGCCGGGGACTGCCCCCTGG  
GCCTGGTGGCGGTGTGCCCCCTTCTCACTGAAGACATGCAGGCCCTGACTCTCCGCACACTGGCCGCCAGCG  
ACGTACCAAGCACTACGAAGTAGTCCGGGAGCTGGGCAAAGGCACCTATGGGAAGGTTGACCTGGTGGTC  
TACAAGGGCACAGGCACAAAAATGGCACTGAAGTTTGTGAACAAGAGCAAAACCAAGCTGAAGAACTTCCT  
ACGGGAGGTGAGCATCACCAACAGCCTCTCCTCCAGCCCCCTTCATCATCAAGGTCTTTGACGTGGTCTTTG  
AGACAGAGGACTGCTACGTCTTTGCCAGGAGTACGCACCTGCTGGGGACCTGTTTGACATCATCCCTCCC  
CAGGTGGGGCTCCCTGAGGACACGGTGAAGCGCTGTGTGCAGCAGCTGGGCCTGGCGCTGGACTTCATGCA  
CGGGCGGCAGCTGGTGCACCGCGACATCAAGCCCCGAGAAGCTGCTGCTGTTTCGACCGCGAGTGGCGCCGCG  
TAAAGCTGGCCGACTTCGGCATGACGCGCCGCGTGGGCTGCCGCGTCAAGCGCGTGAGCGGCACCATCCCT  
TACACGGCGCTGAGGTGTGCCAGGCGGGCCGCGCCGACGGGCTGGCGGTGGACACGGGCGTGGACGTGTG  
GGCCTTCGGCGTGTCTATCTTCTGCGTGTCAACGGCAACTTCCCGTGGGAGGCGGGCGTGGGCGCCGACG  
CCTTCTTCGAGGAGTTCGTGCGCTGGCAGCGGGGCCGCTGCCGGGGCTGCCTTCGCAGTGGCGCCGCTTC  
ACCGAGCCCGCGCTGCGCATGTTCCAGCGCTTACTGGCCCTGGAGCCGAGCGCCGCGGCCAGCCAAGGA  
GGTGTTCGCTTCCCTCAAGCACGAGCTCACGTCCGAGCTGCGCCGCGGCCCTCGCACCGCGCGCGCAAGC  
CCCCGGGGACCGCCCGCCGCGCGGCCACTGCGCCTCGAGGCGCTGGGCGCTCAAGCGGACGGTGTG  
CTGACCGAGAGCGGCGCGGCTCCCGGCCCGCGCCCCCGCGCTCGGGTCGGTGCCCTTGCCCGTGGCGGT  
GCCGTGTCCAGTGGCCGTGCCGTGCTGTGCCCCGAGCCCGGCTAGCTCCCCAGGGGCCCCCGGCCGGA  
CCGACGGCCGCGCGGACAAGAGCAAAGGGCAGGTGGTGTGCGCACGGCCATCGAGATCTGCGTCTGA

>SGK037\_ID#NA\_45

ATGGATAAGTACGATGTGATTAAGGCCATCGGGCAAGGTGCCTTCGGGAAAGCATACTTAGCTAAAGGGAA  
ATCAGATAGCAAGCACTGTGTCTATAAAGAGATCAATTTTGAAAAGATGCCCATACAAGAAAAAGAGCTT  
CAAAGAAAGAGTGATTCTTCTGGAAAAGATGAAACATCCCAACATTGTAGCCTTCTTCAATTCATTTCAA  
GAGAATGGCAGGCTGTTTATTGTAATGGAATATTGTGATGGAGGGGATCTCATGAAAAGGATCAATAGACA  
ACGGGGTGTGTTATTTAGTGAAGATCAGATCCTCGGTTGGTTTGTACAGATTTCTCTAGGACTAAAACATA  
TTCATGACAGGAAGATATTACACAGGGACATAAAAGCTCAGAACATTTTCTTAGCAAGAACGGAATGGTG  
GCAAAGCTTGGGGACTTTGGTATAGCAAGAGTCTGTAATAATTCCATGGAACTTGCTCGAACTTGTATTGG  
AACACCTTACTACCTGTCCCAGAGATCTGTGAGAATAAACCTACAAATAAAACGGATATTGGTCTC  
TTGGCTGTGCTTATATGAGCTCTGCACACTTAAACATCCTTTTGAGGGTAACAATTACAGCAGCTGGTT  
CTGAAGATTGTCAAGCACATTTTGGCCCAATATCTCGGGGTTTCTCGTGAGCTCCATTCCTTGATATC  
TCAGCTCTTCAAGTATCTCTCGAGCCGACCATCCATAAATTCCATTTTGAAAAGGCCCTTTTTAGAGA  
ATCTTATTCCCAAATATTTGACTCCTGAGGTAAGTTTGGAGGTGACTGTTTGGATTTTGGCAGAGATTTG  
GGTTGCAGGTCTTTGACACGTGTGTTTCGGTTTTAGGTGTGAGCCACTGCACCCAGCCTGTATTATGGTTTT  
TAAAAACATCCCCTCTTGTCTTTCAGATAAAAAATGATAGAAAGACCCAAAATTGCTGTCTGTGAGAC  
ATTATGATTATTATTATGCTCAACTTGATATGCTGAGGAGGAGAGCCCAAAACCAAGTTATCACCTATT  
CCTCAAGAAAATACTGGAGTTGAGGATTACGGTCAGGAAACGAGGCATGGTCCATCCCCAAGTCAATGGCC  
TGCTGAGTACCTTCAGAGAAAATTGAAGCTCAACAATATAAGTTGAAAGTGGAGAAGCAATTGGGTCTTC  
GTCCATCTCTGCGGAGCCAAATTACAACCAGAGACAAGAGCTAAGAAGTAATGGAGAAGAGCCTAGATTC  
CAGGAGCTGCCATTTAGGAAAAACGAAATGAAGGAACAGGAATATTGGAAGCAGTTAGAGGAAATACGCCA  
ACAGTACCACAATGACATGAAAGAAATTAGAAAGAAGATGGGGAGAGAACAGAGGACATTGAAAAGACT  
TGAAACAAATGAGGCTTCAGAACACAAAGGAAAGTAAAAATCCAGAACAGAAATATAAAGCTAAGAAGGGG  
GTAAAAATTTGAAATTAATTTAGACAAATGTATTTCTGATGAAAACATCCTCCAAGAGGAAGAGGCAATGGA  
TATACCAATGAACTTTGACCTTTGAGGATGGCATGAAGTTTAAGGAATATGAATGTGTAAAGGAGCATG  
GAGATTATACAGACAAAGCATTGAAAACTTCACTGCCAGAGCAGGGTTTTCCACGCAGACTGTAGCT  
GCTGTGGGAAACAGGAGGAGTGGGATGGAGGAGCGCCTCAGACTCTGCTGCAGATGATGGCAGTGGCCGA  
CATCACCTCCACCTGCCCCACGGGGCCTGACAATGGCCAAGTTATTGTGATTGAAGGCATTCCAGGAAACA

Figure 1X

GGAAACAGTGGCGGCATGAAGCTCCAGGAACCTTAAATGAGTGTTTTGGCAGCAGCACATCTAACGAGTAGC  
TCATTTTCTGCGGATGAAGAATTTGGTATG

>SGK060\_ID#NA\_46

ATGCTGAAATTCCAAGAGGCAGCTAAGTGTGTGAGTGGATCAACAGCCATTTCCACTTATCCAAAGACCTT  
GATTGCAAGAAGATACGTGCTTCAACAAAACTTGGCAGTGGAAAGTTTTGGAACGTCTATCTGGTTTTAG  
ACAAGAAAGCCAAACGAGGAGAGGAATTAAAGGTACTTAAAGGAAATATCTGTTGGAGAACTAAATCCAAAT  
GAAACTGTACAGGCCAATTTGGAAGCCCACTCCTCTCCAAGCTGGACCACCCAGCCATTGTCAAGTTCCA  
TGCAAGTTTTGTGGAGCAAGATAATTTCTGCATTATCACGGAGTACTGTGAGGGCCGAGATCTGGACGATA  
AAATTCAAGGAATATAACAAGCTGGAAAAATCTTTCAGAAAATCAAATAATAGAATGGTTTTATCCAGCTG  
CTGCTGGGAGTTGACTACATGCATGAGAGGAGGATACTTCATCGAGACTTAAAGTCAAAGAATGTATTTCT  
GAAAAATAATCTCCTTAAATTTGGAGATTTTGGAGTTTCTCGACTTCTAATGGGATCCTGTGACCTGGCCA  
CAACTTTAAGTGAACCTCCCATTTATATGAGTCTGAGGCTCTGAAACACCAAGGCTATGACACAAAGTCG  
GACATCTGGTCACTGGCATGCATTTTGTATGAGATGTGCTGCATGAATCATGCATTCGCTGGCTCCAATTT  
CTTATCCATTGTTTTAAAAATTGTTGAAGGTGACACACCTTCTCTCCCTGAGAGATATCCAAAAGAACTAA  
ATGCCATCATGGAAGCATGTTGAACAAGAATCCTTCATTAAGACCATCTGCTATCGAAATTTTAAAAATC  
CCTTACCTTGATGAGCAGCTACAGAACCTAATGTGTAGATATTCAGAAATGACTCTGGAAGACAAAAATTT  
GGATTGTGAGAAGGAGGCTGCTCATATAATTAATGCCATGCAAAAAGGATCCACCTGCAGACTCTGAGGG  
CACTGTGAGAAGTACAGAAAATGACGCCAAGAGAAAGGATGCGGCTGAGGAAGCTCCAGGCGGCTGATGAG  
AAAGCCAGGAAGCTGAAAAAGATTGTGGAAGAAAAATATGAAGAAAATAGCAACGAATGCAAGAATTGAG  
ATCTCGGAACCTTTCAGCAGCTGAGTGTGATGTACTCCATGAAAAACACATTTAAAGGAATGGAAGAAA  
AGGAGGAGCAACCTGAGGGAAGACTTTCTTGTTCACCCAGGACGAGGATGAAGAGAGGTGGCAAGGCAGG  
GAAGAGGAATCTGATGAACCAACTTTAGAGAACCTGCCTGAGTCTCAGCCTATTCTTCCATGGACCTCCA  
CGAACTTGAATCAATTGTAGAGGATGCCACATCTGACCTTGGATACCATGAGATCCCAGAAGACCCACTTG  
TGGCTGAAGAGTACTACGCTGATGCATTTGATTCTTATTGTGTAGAGAGTGTAGAGGAGGAAGAAGAAATA  
GCGTTAGAAAGACCAGAGAAAGAAATCAGGAATGAGGGATCCCAGCCTGCTTACAGAACAAACCAACAGGA  
CAGTGATATCGAAGCGTTGGCCAGGTGTTTGGAAAATGTCTGGGTGCACTTCTCTAGACACAAAGACCA  
TCACCACCATGGCTGAAGACATGTCCCAGGACCACCAATTTTCAACAGTGTGATGGCCAGGACCAAGATG  
AAACGCATGAGGGAATCAGCCATGCAGAAGCTGGGGACAGAAGTATTTGAAGAGGTCTATAATTACCTCAA  
GAGACGAAGGCATCAGAAATGCTAGCGAAGCAGAGATCCGCGAGTGTGTTGAAAAAGTGGTGCCTCAAGCCA  
GCGACTGTTTTGAAGTGGACCAGCTCCTGTACTTTGAAGAGCAGTTGCTGATCAGGATGGGAAAAGAACCT  
ACTCTCCAGAACCATCTCTAG

>SGK080\_ID#NA\_47

ATGCCGTCCCGGGCTGAGAACTATGAAGTGTGTACACCATTTGGCACAGGCTCCTGTGGCCGCTGCCAGAA  
GATCCAGAGGAAGAGTGACGGCAAGATACTAGTTTGGAAAGAACTTCATTATGGCTCCATGACAGAAGCTG  
AGAAACAGATGCTTGTCTGAAAGTGAATTTGCTTGTGAACTGAAAAATCCAAACATCGTTTATTACTAT  
GATCGTATTATTGACCGACCAACACAACACTGTACATTGTAATGGAATATTGTGAAGAAGGAGACCTGGC  
TAGTGTAATTACAAAGGGAACCAAGGAAAGGCAATACTTAGATGAAGAGTTTGTCTTCGAGTGACAACCTC  
AGTTGACTCTGGCCCTGAAGTNATGCCACAGACGAAGTGATGGTGATCATACTGTAGTGCGTCCGGATCTG  
AAACCAGCCAGTGTTCCTGGATGGCAAGCAAAACGTCAAGCTTGGAGATTTGGGGCTAGCCAGAATATT  
AAACCAGACACGAGTTTTCGAAAAACATTTGTTGGCACACCTTATTACATGTCTCCTGAACAAACGAATC  
ACATGTCTTACAATGAGAAACAGATATCTGGTCATTGGGCTGCTTGCCTATGAGTCACGTGCATTAATG  
CCTCCATTTACAGCTTTTAGCCAGAAAGAACTCGCTGGGAAAATCAGAGAAGGCAAAATTCAGGCGAATTCT  
ATACCGTTACTCTGATGAATTGAATGAAATTATTATGAGGATGTAAAGGATTACCATCGACCTTCTGTTG  
AAGAAATTCTCGAGAACCCTTTAATAGCAGATTTGGTTGCGAGAAGAGCAAGAAGAAATCTTGAGAGAAGA  
GGGCGACAATTAGGAGAGCCAGAAAAATTGCCGGATTCAGCCCTGTATTGAGTGAGCTGAAACTAAAGGA  
AATTCAGTTAGAGGAGCAAGAGCGAGCTCTCAAAGCAGGAGAAGAAAGGTTGGAGCAGAAAGAACAGGAGC  
TTTGTGTTTGTGAGAGACTAGCAGAGGACAGACTGGCTATACCAGAAAATCTGTTGAAGAATTACAGCTTG  
CTAAAGGAACAGAAGTTCCTGTCTCTGGCAAGTAGTCCAGAACTTCTTAATCTTCCATCCTCAGTAATTAA  
GAAGAAAGTTTCAATTCAGTGGGGAAAGTAAAGAGAATGTGATGAGGAGTGAGAATCCTGAGAGTCAGCTCA  
CATCTAAGTCCAAGTGCAAGGACCTGAAGNNNNNTGCTTCATGCTTCATGCTGCCAGCTGCGGGCTCAA  
GCCCTGTGATATTTAGAAAAATTATCAACTAAAAAGCAGACAGATCCTGGGCATGCGC

Figure 1Y

>SGK002\_ID#NA\_48  
TAATAATAGTACCTACCTCATGTTCATCGTTTTAAGGATTAAATGAATTAACACATTTAAATATAAGAAGT  
TTGAAAAGCTCTTGGACTCCGGCTGCAGCGTCAGCCTCCCTCGGGCCTCGGCAGCGGCGGGCTCGCTCG  
CCTCAGCCCCAGCGCCCCCTCGGCTACCTTCGGACCAGGCCCCGAGCGCCCCCGCCCTCGGCCGCCCGAC  
GCCGGCCTGGGCGCGGTTCGAGCCCCGGGCTCGCGTAGGCGCCGAACGCTCCCGGCCCGCCACTACGGG  
CCCAGGCTAGAGCGCGCGCCGACCGGCCCGGAGCCCGGATGCTGGCCCGGAGGAAGCCGATGCTGCC  
GGCGCTCACCATCAACCCTACCATCGCGAGGGCCCCGTCCCCAACAGCGAGGGCGCCTCCGAGGCAAACC  
TGGTGGACCTGCAGAAGAAGCTGGAGGAGCTGGAACCTTGACGAGCAGCAGAAGCGGCTGGAAGCCTTTCTC  
ACCCAGAAAGCCAAGGTCCGGCGAACTCAAAGACGATGACTTCGAAAGGACCTCAGAGCTGGACGCGGGCAA  
CGGCGGGGTGGTCACCAAAGTCCAGCACAGACCCTCGGGCCTCATCATGGCCAGGAAGCTGATCCACCTTG  
AGATCAAGCCGGCCATCCGGAACCAGATCATCCGCGAGCACCAGGTCTGACGAGTGAACCTACCGTAC  
ATCGTGGGCTTCTACGGGGCCTTCTACTGTGACAGGGAGATCAGCATCTGCATGGAGCACATGGATGGCGG  
CTCCCTGGACCAGGGGCTGAAAGAGGCCAAGAGGATTCCCGAGGACATCCTGGGGAAGTCAAGCTTGCAG  
TTCTCCGGGGCTTGGCGTACCTCCGAGAGAAGCACCAGATCATGCACCGAAATGTGAAGCCCTCCAACATC  
CTCGTGAACCTTAGAGGGGAGATCAAGCTGTGTGACTTCGGGGTGAGCGGCCAGCTCATCGACTCCATGGC  
CAACTCCTTCGTGGGCACGCGCTCCTACATGGCTCCGGAGCGGTTGACAGGGCACACATTACTCGGTGACGT  
CGGTCTATCTGGAGCATGGACCTGTCCCTGGTGGAGCTGGCCATCGAAAGGTACCCCATCCCCCGCCGAC  
GCCAAGGAGCTGGAGGCCATCTTTGGCCAGCCCGTGGTTCGACAGGGAAGAAGGAGAGCCTCACAGCATCTC  
CTCTTGGCCAGGGTCCCCCGGGCGCCCCAACAGCGGTTACGGGATGGACAGCCTGCCCGCCATGGCCATCT  
TCGAACTGCTGGACTATATTGTGAAGAGCCGCTCCTAAGCTGCCCAACGGTGTGTACACCCCGGCTTC  
CAGGAGTTTGTCAATAAATGCCTCATCAAAAACCCACGGAGCGGGCGGACCTAAAGATGCTCAAAACCA  
CGCCTTCATCAAGCGGTCCGAGGTGAAAGAAGCGGATTTTGCTGCNNNTTGTGTAAAACCTGGGGCTGA  
ACCAGCCCGGCACACCCACGCGCACCGCCGTGTGACCGGCACACCCACGCGCACCGCCGTGTAACAGTGGC  
CAGGCTCCCCCGCTCCCGCTGGTGACCTGCCACCGTCCCTGTCCACGCCCCGCGCTTCCAGCTGAGGACA  
GGCTGGCGCCTCCACCCACCTCCTGCCTCACCCCTGCGGAGAGCACCGTGGCGGGGCGACAAGTGGCAC  
GCAGGAACGGGGGTCTCCTCTCCTGCTGGTCCCTGCCGCGTGCCTCTAGGGACGGGCGACGCTGCTGTGT  
TGCTCTCAGAGCCTCTGCTTCCTTAGGTTACAAAACAAAACAGGGAGAGAAAA

>SGK058\_ID#NA\_49  
AATAGTTTAAAGTCTGAAGAACCTATCCTATGGACCAAGGGTGAGATTCTTGGAAAGGGAGCCTACGGCAC  
AGTATACTGTGGTCTCACTAGTCAAGGACAGCTAATAGCTGTAAACAGGTGGCTTTGGATACCTCTAATA  
AATTAGCTGCTGAAAAGGAATACCGGAAACTACAGGAAGAAGTAGATTGCTCAAAGCACTGAAACATGTC  
AACATTGTGGCCTATTTGGGGACATGCTTGCAAGAGAACACTGTGAGCATTTCATGGAGTTTGTTCCTGG  
TGGCTCAATCTCTAGTATTATAAACCGTTTGGGCCATTGCCTGAGATGGTGTCTGTAAATATACGAAAC  
AAATACTTCAAGGTGTTGCTTATCTCCATGAGAACTGTGTGGTACATCGCGATATCAAAGGAAATAATGTT  
ATGCTCATGCCAACTGGAATAATAAGCTGATTGACTTTGGCTGTGCCAGGCGTTTGGCTGGGCAAGTTT  
AAATGGCACCCACAGTGACATGCTTAAGTCCATGCATGGGACTCCATATTGGATGGCCCCAGAAGTCATCA  
ATGAGTCTGGCTATGGACGGAATCAGATATCTGGAGCATTGGTTGTACTGTGTTTGGATGGCTACAGGG  
AAGCCTCCACTGGCTTCCATGGACAGGATGGCCGCCATGTTTTACATCGGAGCACACCGAGGGCTGATGCC  
TCCTTTACCAGACCACTTCTCAGAAAATGCAGCAGACTTTGTGCGCATGTGCCTGACCAGGGACCGCATG  
AGCGACCTTCTGCTCTCCAGCTCCTGAAGCACTCCTTCTTGGAGAGAAGTCAC

>SGK103\_ID#NA\_50  
GATCTAACCCTCAGAATATCCTGCTTGACAAGAGTGTCTTTGTATACCTGGGGGTGTTTGGCCTTGCCAG  
AGAGTTTATGCTA

>SGK035\_ID#NA\_51  
AGAGACTGCCCTCCCGTTATTGCCCCACCACCGGATCATATGAAATCAATTTACACACGGGCTGTAATTGA  
TCCTGTTCTTGACACAGTTGGTGATTCAAATGTTGATGGTGGTGCCAAGTCTTTAGACAAACAGAAAAAGA  
AGACTAAGATGACAGATGAAGAGATTATGGAGAACTAAGAAGTATGTGAGCATAGGACCCGTAAGAAAA  
AAATATACAAGATATGAAAAAATTGGACAAGGGGCTTCTGGTACAGTTTTCACTGCTACCGACGTTGCACT  
GGGACAGAAGGTTGCTATCAAACAAATTAATTTACAGAAACAGCCAAAGAAGGAATTGATCATTAAATGAGA

Figure 12

TTCTGGTAATGAAAGAATTAAAAAATCCCAACATAGTTAACTTCTTGGACAGTTACCTGGTAGGAGATGAA  
TTGTTTGTGGTCGTGGAATACCTTGCTAGGGGGTCACTCACTGATGTGGTAACAGAAACCTGCATGGATGA  
AGCACAGATTGCTGCTGTATGCAGAGAGAGTTTACAGGCATTGGAGTTTTTACATGCTAATCAAGTGATCC  
ACAGAGACATCAAAAGTGACAGTGTACTTTTGGGAATGGAAGGATCAGTTAAGCTCACTGACTTTGGTTTC  
TGTGCCCAGATCACCCCTGAGCAGAGCAAACGCAGTACCGTGGTCAGAACGCCCACTGGATGGCACCAGA  
AGTGGTTACACGGAAGGCTTATGGCCCTAAAGTCAGTGTATGGTCTCTGGGTATCATGGCTACTGAGATGG  
TAGAAGGAGAGCCTCCATACCTCAATGAAAAATCCCTTGAGGGCCTTGTGCCTAATAGCAACTAATGGAATC  
CCAGAATTCAGAATCCAGAGACACTTTCCCAATATTTTCGGGATTTCTTAAATCGATGTTTGGAAACAGA  
TGTGGAAAAAGGGGTTTCCAGCCAAAGAATTATTACAGCATCTTTTCTGAAACTAGCCAAACTGTTATCTA  
GCTTGACACCACTGATCATGGCAGCTAAAGAAGCAATGAAGAGTAACCGT

>SGK075\_ID#NA\_52

ATGATGCTTGGACTTGAATCACTTCCAGATCCCACAGACACCTGGGAAATTATAGAGACCATTGGTAAAGG  
CACCTATGGCAAAGTCTACAAGGTAACAAAGAGAGATGGGAGCCTGGCTGCAGTGAATTTCTGGATC  
CAGTCAGTGATATGGATGAAGAAATTGAGGCAGAAATACAACATTTTGCAGTTCTTCTTAATCATCCCAAT  
GTTGTAAAGTTTTATGGGATGTTTTACAAAGCGGATCACTGTGTAGGGGACAGCTGTGGCTGGTCTCTGGA  
GCTGTGTAATGGGGGCTCAGTCACTGAGCTTGTCAAAGGTCTACTCAGATGTGGCCAGCGGTGGATGAAG  
CAATGATCTCATACATCTTGTACGGGGCCCTCTTGGGCCCTTCAGCATTTGCAACAACCGAATCATCCAC  
CGTGATGTGAAGGGGAATAACATTCTTCTGACAACAGAAGGAGGAGTTAAGCTCGTTGACTTTGGTGTTC  
AGCTCAACTCACCAGTACACGTCTGCGGAGAAAACATCTGTTGGCACCCCGTTCTGGATGGCCCTGAGG  
TCATTGCCTGTGAGCAGCAGTATGACTCTTCTATGACGCTCGCTGTGACGTCTGGTCTTGGGGATCACA  
GCTATTGAACTGGGGGATGGAGACCTCCCTCTTTGACATGCATCCTGTGAAAACTCTTTAAGATTCC  
AAGAAATCCTCCACCTACTTTACTTTCATCCAGAAAAATGGTGTGAAGAATTCAACCACTTTATTTTACAGT  
GTCTTATTAAGGATTTTGAAAGGCGACCTTCCGTCACACATCTCCTTGACCACCCATTTATTAAGGAGTA  
CATGGAAGGTTCTGTTTCTGCAAAAACAGCTGGCCAAGGTTCTCCAAGACCAGAAGCATCAAAATCCTGT  
TGCTAAAACCAGGCATGAGAGGATGCATACCAGAAGACCTTATCATGTGGAAGATGCTGAAAAATACTGCC  
TTGAGGATGATTTGGTCAACCTAGAGGTTCTGGATGAGGATACAATTATCCATCAGTTGCAGAAACGTTAT  
GCAGACTTGCTAATTTACACATATGTTGGAGACATCTTAATTGCCTTAAACCCCTTCCAGAATCTAAGCAT  
ATACTCTCCACAGTTTCCAGACTTTATCATGGGGTGAACGCGCCTCCAATCCCCCCCACATATTTGCAT  
CAGCAGATGCTGCTTACCAGTGCATGGTTACTCTCAGCAAAGACCAGTGCATTGTGCATCAGCGGAGAGAGT  
GGCTCTGGGAAGACAGAAAGCGCCACCTGATTTGTTTTCAGCATTTGACTTTCTTGGGAAAGGCCAATAATCA  
GACCTTGAGAGAGAAAATCTACAAGTCAACTCCCTGGTGAAGCCTTTGGGAACTCATGCACTGCCATCA  
ATGACAACCTCGAGCCGTTTGGAAAAATATCTGGAAATGATGTTTACACCACTGGAGTTGTGGCGGGCA  
AGAATCTCTGAATATCTCCTGGAATAATCCAGAGTTATAAAACAGGCAGCGAGAGAGAAAAATTTTCATAT  
ATTTTACTATATTTATGCTGGTCTTCATCACCAAAAGAAGCTTTCTGATTTTCACTTCTCTGAGGAAAAAC  
CTCCTAGGTACATAGCTGATGAACTGGAAGGGTGATGCACGACATAACTTCCAAGGAGTCTTACAGAAGA  
CAATTGCAAGCAATTCAGCATTGCTTCAGGATTATAGGGTTACGGACAAAGAGGTGCACTCAGTGTACAG  
AATTTTGGCTGGGATTTTGAATATTGGGAACATTGAGTTTCGAGCTATTTCTCTCAACATCAGACTGATA  
AAAGTGAGGTGCCCAATGCTGAAGCTTTGCAAAATGCTGCCTCTGTTCTGTGCATTAGCCCTGAAGAGCTC  
CAGGAGGCCCTCACCTCCCACTGTGTGGTCAACCGGGGCGAGACCATCATCCGTGCCAACACTGTAGACAG  
GGCTGCGGACGTTTCGAGACGCCATGTCCAAAGCCCTGTATGGGAGGCTCTTCAGCTGGATTGTGAATCGCA  
TTAATACACTCCTGCAAGCCAGACGAAAACATATGGCAAGTTCTTCGGAGAGCAGAGGGAGGTGGAATGAAT  
GTGGGGATCTTGGATATCTTTGGATTTCGAGAATTTTTCAGAGAAATTCATTGAGCAGCTCTGCATAAACAT  
CGCCAATGAGCAAATCCAGTACTATTTCAATCAGCATGTTTTTGTCTCTTGGAGCAGATGGAATATCAGAATG  
AAGGCATTGATGCTATACCCGTGGAATATGAGGACAACCGCCCGCTCTCGGACATGTTCTCCAGAAACCC  
CTGGGAGTGCCTGCACTTTTGGATGAGGAAAGTCGGTTTCCCAAGCAACTGACCAGACCCTGGTTGATAA  
ATTTGAAGATAATCTACGATGCAAAATACTTCTGGAGGCCCAAAGGAGTGGAACTGTGCTTTGGCATTTCAG  
ATTATGCTGGAAAGGTATTATATGATGCTTCTGGGGTCTTGTGAGAAAAATAGAGACACTCTCCCTGCCGAT  
GTGGTTGTGGTCTCTGAGAAGCTCAGAAAACAAGCTTCTCAGCAGCTCTTCTCAATCCCTCTGACCAAAAC  
AGGTACTTGGGAACCCCTCTGATAGCCCTGCTTAAAGCTTTTGAAGACCAGGTAATTTGGCCCGAGACAA  
GAGCTAGGATAACAGTGGCCTCAAGTTCTTTCCTCCACATTTTCAGTGTGGGAAAGCCAAGGTGGACACT  
CTGGAGGTGATACGGCATCCGGAAGAAACCACCAACATGAAGAGGCAAACTGTGGCTTCTTACTTCCGGTA  
TTCTCTGATGGACCTGCTCTCCAAATGGTGGTTGGACAGCCCCACTTTGTGCGCTGCATTAAACCCAATG

Figure 1AA



ATGACCGAGAGGCCCTGCAGTTCTCTCGAGAGAGGGTGCTGGCCAGCTCCGCTCCACAGGGATTCTGGAG  
ACAGTCAGCATCCGCCGCCAGGGCTATTCCCACCGCATCCTTTTTGAAGAATTTGTGAAAAGGTATTATTA  
CTTGGCATTACAGCACATCAAACACCTCTTGCTAGCAAAGAGAGCTGTGTGGCTATCTTGAAAAGTCCA  
GATTAGATCACTGGGTGCTGGGAAAAACAAAGGTTTTTCTCAAATATTACCATGTTGAGCAATTAAATTTG  
CTGCTTCGAGAAGTCATAGGCAGAGTGTTGTGTGTCAGGCATATACCAAGGGGTGGCTTGAGCCAGGAG  
ATACAAAAGGTCAGAGAGAAGAGAGAGAAGGGAGCCATTGCCATCCAGTCA

>SGK188\_EPHA9\_ID#NA\_53

GGGGGCATTGCTCAGCGGTGCTAGGCTGGCGCGGCTTGAGCCGCCGCCGGAAGTGCAGCTCGGTCTGCGGA  
CCATGGAGACCTGCGCCGGTCCACACCCGCTGCGCCTCTTCTCTGCCGGATGCAGCTCTGTCTCGCGCTG  
CTTTTGGGACCCTGGCGGCCTGGGACCGCCGAGGAAGTTATCCTCCTGGATTCCAAAGCCTCCCAGGCCGA  
GCTGGGCTGGACTGCACTGCCAAGTAATGGGTGGGAGGAGATCAGCGCGTGGATGAACACGACCGTCCCA  
TCCGCACGTACCAAGTGTGCAATGTGCTGGAGCCCAACCAGGACAACCTGGCTGCAGACTGGCTGGATAAGC  
CGTGGCCCGCGGCAGCGCATCTTCGTGGAAGTGCAGTTCACACTCCGTGACTGCAGCAGCATCCCTGGCGC  
CGCGGTACCTGCAAGGAGACCTTCAACGTCTACTACCTGGAACTGAGGCCGACCTGGGCCGTGGGCGTC  
CCCGCTAGGCGGCAGCCGCCCGCAAAATCGACACGATCGCGCGGACGAGAGCTTCACGACGGGCGAC  
CTGGGTGAGCGCAAGATGAAGCTGAACACAGAGGTGCGCGAGATCGGACCGCTCAGCCGCGGGGTTTCCA  
CCTGGCCTTTTTCAGGACGTGGGCGCATGCGTGGCGCTTGTCTCGGTGCGCGTCTACTACAAGCAGTGCCGCG  
CCACCGTGGCGGGCCTGGCCACGTTCACAGCCACCGCAGCCGAGAGCGCCTTCTCCCACTGGTGGAAAGTG  
GCCGGAACGTGCGTGGCGCACTCGGAAGGGGAGCCTGGCAGCCCCCAGCATGCACTGCGGCGCCGACGG  
CGAGTGGCTGGTGCCTGTGGGCGCTGCAGCTGCAGCGCGGATTCCAGGAGCGTGGTGACATCTGCGAAG  
CCTGTCCCCCAGGGTTTTACAAGGTGTCCCCGCGCGAAGGGTCTGCTCACCGTGCCAGAGCACAGCCGG  
GCCCTGGAACACGCTCCACCTTCTGCGTGTGCCAGGACAGCTATGCGCGCTCACCCACCGACCCGCTC  
GGCTTCTGTCACCCGTGGGCGCGCGTGGCGCGCGGACCTGCAGTACAGCCTGAGCCGCTCGCGCGCTGG  
TGCTGCGACTGCGCTGGCTGCCGCGCGGCGACTCGGGAGGCGCGCTCGGACGTCACCTACTCGCTGCTGTGC  
CTGCGCTGCGGCCGCGAGGGCCCCGCGGGCGCCTGCGAGCCGTGCGGGCGCGCGTGGCCTTCTACCGCG  
CCAGGCAGGGCTGCGGGAGCGAGCCGCCACGCTGCTGCACCTGCGGCCCGGGGCGCGCTACACCGTGC  
TGGCGGTGCTCAACGGCGTCTCGGGCCCCGCGCGCGCCCTGGTTCCGGTTGGCGCTGTTTCAATTAACCT  
GGTACGGTTGGCCCTGTTCTGTTGCCGGGTTATCCGCGACCGAGTGAACCCAGAGCGTGTCCCTGTC  
GTGGCGGGAGCCCATCCCTGCCGAGCCCCGTGGGGCAATGACACGGAGTACGAGATCCGATACTACGAGA  
AGGTGCAGAGTGAGCAGACTTACTCCATGGTGAAGACAGGGGCGCCACAGTCAACCGTCAACCAACCTGAAG  
CCGGCTACCCGCTACGTCTTTTTCAGATCCGGGCGCCTTCCCCGGGGCCATCCTGGGAGGCCCAGAGTTTTAA  
CCCCAGCATTGAAGTACAGACCTGGGGGAGGCTGCCTCAGGGTCCAGGGACAGAGCCCCGCCATTGTGCG  
TCACCGTAGTGACCATCTCGGCCCTCCTCGTCTGGGCTCCGTGATGAGTGTGCTGGCCATTTGGAGGAGG  
CCCTGCAGCTATGGCAAAGGAGGAGGGGATGCCCATGATGAAGAGGAGCTGTATTTCCACTTCAAAGTCCC  
AACACGTGCGACATTCTGGAACCCAGAGCTGTGGGGACCTGCTGCAGGCTGTGCATCTGTTCCGCAAGG  
AACTGGATGCGAAAAGCGTCACGCTGGAGAGGAGCCTTGGAGGAGGGCGGTTTGGGGAGCTGTGCTGTGGC  
TGCTTGCAGCTCCCGGTGCGCAGGAGCTGCTCGTAGCCGTGCACATGCTGAGGGACAGCGCCTCCGACTC  
ACAGAGGCTCGGCTTCTTGGCCGAGGCCCTCAGCTGGGCGAGTTTGACCATAGCCACATCGTGCAGGCTGG  
AGGGCGTTGTTACCCGAGGAAGCACCTTGATGATTGTACCGAGTACATGAGCCATGGGGCCCTGGACGCG  
TTCTCAGGCGGCACGAGGGGAGCTGGTGGCTGGGCAACTGATGGGGTTGCTGCTGGGCTGGCATCAGC  
CATGAAGTATCTGTGAGAGATGGGCTACGTTACCGGGGCGCTGGCAGCTCGCCATGTGCTGGTGCAGCAGC  
ACCTTGTCTGCAAGATCTCTGGCTTGGGCGGGGCCCCCGGGACCGATCAGAGGCTGTCTACACCACTATG  
AGTGGCCGGAGCCCAGCGCTATGGGCCGCTCCCGAGACACTTCAGTTTGGCCACTTCAGCTCTGCCAGTGA  
CGTGTGGAGCTTCGGCATCATATGTGGGAGGTGATGGCCTTTGGGGAGCGGCGCTTACTGGGACATGTCTG  
GCCAAGACGTGATCAAGGCTGTGGAGGATGGCTTCCGGCTGCCACCCCCAGGAAGTGTCTTAACCTTCTG  
CACCGCTAATGCTCGACTGCTGGCAGAAGGACCCAGGTGAGCGGCCCAGGTTCTCCAGATCCACAGCAT  
CCTGAGCAAGATGGTGCAGGACCCAGAGCCCCCAAGTGTGCCCTGACTACCTGTCCAGGCTCCCAACCC  
CACTAGCGGACCGTGCCTTCTCACCTTCCCCCTCTTTGGCTCTGTGGGCGCGTGGCTGGAGGCCCTGGAC  
CTGTGCCGCTACAAGGACAGCTTTCGCGGCTGCTGGCTATGGGAGCCTGGAGGCCGTGGCCGAGATGACTGC  
CCAGAGGGACCTGGTGAAGCTAGGCATCTCTTGGCTGAACATCGAGAGGCCCTCTCAGCGGGATCAGCG  
CCCTGCAGGCACGAGTGTCCAGCTGCAGGGCCAGGGGTGCAGGTGTGAGTGGACCCC

Figure 1BB



&gt;SGK040\_ID#NA\_54

ATGCAGCCCCCTGGACTTTAGTTTCAGGGGGAAGTGACCCCAACATCAGCCTCTCAGAAAAAGATCCGAGATCA  
GCTTGTGTGGACAGCTGATTCCAGACTGCTATGTAGAAGTTGAAAAATCATTATTCGGAGCGTAAAA  
ATGTGCCAATTGAATTTCCCGTAATTGACCGGAAACGATTATTACAAGTAGTGAGAGAAAAATCAGCTGCAG  
TTAGATGAAAATGAGCTTCTCACGCAGTTCACTTTCTAAATGAATCAGGAGTCCTTCTTCATTTTCAAGA  
CCCAGCACTGCAGTTAAGTGACTTGTACTTTGTGGAACCCAAGTGGCTTTGTAAAAATCATGGCACAGGATG  
TTAGCAGCATTTTTGGCCTTTATATTCGAGACATTTTGACAGTGAAAGTGGAAAGGTGTCCAAAACACCCT  
AAGGGCATTATTTTCGCGTAGAGATGTGAAAAATTTCTTTCAAAAAAAGGAAATTTCCAAAGAACTACAT  
GTCACAGTATTTTAAGCTCCTAGAAAAATTCAGATTGCTTTGCCAATAGGAGAAGAATATTTGCTGGTTC  
CAAGCAGTTTGTCTGACCACAGGCCTGTGATAGAGCTTCCCATTTGTGAGAACTCTGAAATTATCATCCGA  
CTATATGAAATGCCTTATTTTCCAATGGGATTTTGGTCAAGATTAATCAATCGATTACTTGAGATTTACC  
TTACATGCTTTCAGGGAGAGGCTGTATTCTTTTGGGCCAAGTTGTGGACCACATTGATTCTCTCATGGAAG  
AATGGTTTCTGGGTTGCTGGAGATTGATATTTGTGGTGAAGGAGAACTCTGTTGAAGAAATGGGCATTA  
TATAGTTTAAATGATGGTGAAGAACATCAAAAAATCTTACTTGATGACTTGATGAAGAAAGCAGAGGAAGG  
AGATCTCTTAGTAAATCCAGATCAACCAAGGCTCACCATTCCAATATCTCAGATTGCCCTGACTTGATTT  
TGGCTGACCCGCTAGAAATATTATGTTGAATAATGATGAGTTGGAATTTGAACAAGCTCCAGAGTTTCTC  
CTAGACTGTTTGTGTGATTCACTTATATCCATCAAGTGACTACATTTCAAGGCACTATATGAGAACCAT  
AAATATTGTACAAACAGGATTTGCTAAATGTCCGTGGAGAGTAACAGTCCACGGGGCTGATCATGGTGTAG  
GCAGTTTGGATCAGTTTACCGAGCAGCCTATGAAGGAGAAGAAGTGGCTGTGAAGATTTTAAATAACAT  
ACATCACTCAGGCTGTTAAGACAAGAGCTTGTGGTGCTTTGCCACCTCCACCACCCAGTTTGTATATCTTT  
GCTGGCAGCTGGGATTTCGTCCCCGGATGTTGGTGTAGGAGTTAGCCTCCAAGGGTTCTTGGATCGCCTGC  
TTCAGCAGGACAAAAGCCAGCCTCACTAGAACCCTACAGCACAGGATTGCACTCCACGTAGCTGATGGTTTG  
AGATACCTCCACTCAGCCATGATTATATACCGAGACCTGAAACCCACAATGTGCTGCTTTTACACTGTA  
TCCAATGCTGCCATCATTGCAAAGATTGCTGACTACGGCATTGCTCAGTACTGCTGTAGAATGGGGATAA  
AAACATCAGAGGGCACACCAGGGTTTCGTGCACCTGAAGTTGCCAGAGGAAATGTCATTTATAACCAACAG  
GCTGATGTTTATTCATTTGGTTTACTACTCTATGACATTTTGACAAGTGGAGGTAGAATAGTAGAGGGTTT  
GAAGTTTCCAAATGAGTTTGATGAATTAGAAATACAAGGAAATTAACCTGATCCAGTTAAAGAATATGGTT  
GTGCCCCATGGCCTATGGTTGAGAAATTAATTAACAGTGTTTGAAGAAATCCTCAAGAAAGGCCTACT  
TCTGCCCAGGTCTTTGACATTTTGAATTCAGCTGAATTAGTCTGTCTGACGAGACGCATTTTATTACCTAA  
AAACGTAATTGTTGAATGCATGGTTGTACACATCACAAACAGCAGGAATGCAAGCATTGCGCTGGGCTGTG  
GGCACACCGACAGAGGACAGCTCTCATTTCTTGACTTAAATACTGAAGGATACACTTCTGAGGAAGTTGCT  
GATAGTAGAATATTGTGCTTAGCCTTGGTGATCTTCTGTTGAAAAGGAAAGCTGGATTGTGTCTGGGAC  
ACAGTCTGGTACTCTCCTGGTCATCAATACCGAAGATGGGAAAAAGAGACATACCCTAGAAAAGATGACTG  
ATTCTGTCACTTGTGTTGATTGCAATTCCTTTTCCAAGCAAGCAACAAAAAATTTCTTTTGGTTGGA  
ACCGTGTAGGCAAGTTAGCAATTTTGAAGATAAGACTGTTAAGCTTAAAGGAGCTGCTCCTTTGAAGAT  
ACTAAATATAGGAAATGTGAGTACTCCATTGATGTGTTTGAAGTGAATCCACAAAATCAACGGAAAGAAATG  
TAATGTGGGGAGGATGTGGCACAAGATTTTCTCCTTTTCTAATGATTTCAACATTTCAGAAACTCATTGAG  
ACAAGAACAAGCCAAGTGTTCCTTATGCAGCTTTTCAGTGATTCCAACATCATAACAGTGGTGGTAGACAC  
TGCTCTCTATATTGCTAAGCAAAATAGCCCTGTTGTGGAAGTGTGGGATAAGAAAAGTGAAGAACTCTGTG  
GACTAATAGACTGCGTGCATTTTTTAAGGTAA

&gt;SGK390\_ID#NA\_55

ATGGCAGGGGCCGAGGCCAGCACCACCTCCGGGCGCCGCTGGAGGAGCGGCCGCGGAGCCGGCGCCGC  
GGTCACCTCCGCGCTGCCTCGGCGGGGCGGGAGAGGATTCTGCTGACAGCGAAGCGGAGCAAGAGGGAC  
CCCAGAACTGATCCGCAAAGTGTCTACCTCGGGGCGAGATCCGGACCAAGACCAGTATTAAAGAGGGACAG  
CTATTGAAGCAAACCAAGTTCTTTCCAAAGGTGGAAGAAAGCGATACTTCAAACCTTCGAGGCCGACCTTTA  
CTATGCAAAGGACTCAAAGTCTCTGATATTTGATGAAGTTGACCTCTCAGATGCTAGTGTAGCTGAAGCAA  
GCACGAAAAATGCTAACAACAGCTTCACGATCATCACTCCATTTCAGAAAGGCTAATGCTGTGTGCTGAGAAC  
AGAAAGGAGATGGAGGATTGGATCAGCTCACTGAAGTCTGTACAGACCAGAGAACCTACGAGGTGGCCCA  
GTTTAAATGTGGAACATTTCTCAGGGATGCACAACTGGTACGCTGCTCCCACGCCCGACCCACCTTCTGTA  
ACGTTGTGCAGAGAGAGTCTTTCTGAGTCACTCCCATTGGCCTGTCTGCGAAGTGTGTAAATTCAAGGCT  
CAGAAAAGATGTGAGTGAGAGCAACAAATACTGTAAATGGACTACCTGGCCTCCATCGGGAAGGACAT  
TATAGAAGATGAAGATGGCGTCGCGATGCCTCACCAGTGGCTTGAGGGCAACCTGCCTGTAAGTGCCAAGT

Figure 1CC

GTGCTGTCTGCGACAAAACATGTGGCAGTGTTCTCCGTCTACAGGATTGGAAATGCCTTTGGTGTAAGACA  
ATGGTACACACTGCCTGCAAAGATTTATACCATCCAATATGTCCACTTGGTCAATGTAAAGTATCTATCAT  
ACCTCCAATTGCACTAAACAGCACCGATTCCGATGGTTTCTGTAGAGCAACATTTTCGTCTGTGTTAGTC  
CTCTATTGGTTTTTGTCAATTCTAAGAGTGGAGATAATCAGGGAGTAAAGTTCCCTCCGTCCGCTTTAAACAG  
TTGCTAAATCCGGCTCAGGTGTTGATTTAATGAATGGAGGTCCTCATTAGGTTTAAAGATTATTTTCAGAA  
GTTTGACAATTTCCGGATTCTGTATGTGGAGGCGATGGAAGTGTAGGTTGGGTTTTGTGAGAAATCGATA  
AGCTCAACTTGAATAAACAGTGTGAGCTGGGAGTGTGCTTTGGGTACAGGAAATGACCTTGCCCGAGTT  
CTTGCTGGGGAGGTTCAATATGACGATGACACCCAGCTTCCTCAGATCCTAGAGAACTGGAACGAGCCAG  
TACCAAATGTTGGACAGGTGGAGTATAATGACATATGAACCTCAAATGGCCACCAAAAGCTTCCCTACTTC  
CAGGACCTCCAGAAGCATCTGAAGAATTTTATATGACGATTTATGAAGACTCAGTTGCAACGCATCTTACA  
AAAATCCTCAATTCTGATGAACATGCAGTGGTCATATCTTCTGCCAAGACGCTATGTGAACTGTAAAGGA  
CTTCGTTGCCAAAGTAGAAAAGACGTATGACAAAACCTTGGAATGCCGTTGTAGCTGATGCCGTGGCCA  
GTAAATGTTTCAGTCTTAAACGAGAAGCTCGAACAACCTGCTGCAGGCTTTGCACACAGATTCCCAGGCTGCG  
CCTGTTCTCCCTGGCCTCAGCCCTCTCATTGTGGAAGAAGATGCTGTGGAATCGTCCAGTGAAGAGTCCCT  
GGGTGAAAGCAAAGAGCAGCTTGGGGATGACGTTACAAAACCTTCCTCCAGAAAGCCGTCAAACCAAGGG  
AAATCATGTTGCGGCAAAATAGTTTAAAGAAAGCAGTGAGGCAAGTCATTGAGGAAGCCGGAAGTTATG  
GATGACCCGACAGTTCAACCCCTGTGAACAGCTTAATCAGTCTCTGATTATGACAGCACAGAAACAGATGA  
ATCTAAGGAGGAAGCTAAAGATGATGGTGCCAAAGAATCAATAACTGTTAAACTGCACCTCGGTCTCCAG  
ATGCCCGGGCAAGTTATGGCCATTCCCAAACCTGATTCTGTCCCTGGTCCAGCTGTGGCAGCCAGCAAAGAA  
AACCTCCCTGTGCTCAATACCAGAATAATCTGCCAGGTTTAAAGAGCAGGACTGGCTGAGCTCAATTGCTGG  
GAGTTCGATTATCAACAAAATGTTACTGGCAAACATTGATCCTTTTGGTGCCACGCCGTTTATTGACCCCTG  
ATCTAGATTCCGTAGATGGATATTAGAAAAATGTGTATGAACAATTACTTTGGGATTGGATTAGATGCA  
AAAATTTTCATTAGAATTTAATAATAAAAGAGAGGAGCACCTGAAAAATGCAGGAGCCGAACCTAAAAACTT  
GATGTGGTATGGAGTCCCTTGAACCCGGGAGTTATTACAGAGATCGTACAAGAATTTAGAACAAAGGGTTC  
AACTTGAGTGTGATGGGCAGTATATTCTCTTCCAGCTTGCAAGGCATAGCCGTGTTGAACATTCCCAGC  
TATGCTGGAGGCACTAACTTTTGGGGTGGAACTAAAGAGGATGATATATTTGCTGCACCATCCTTTGATGA  
CAAGATCCTGGAAGTTGTAGCAATATTTGATAGCATGCAATGGCAGTTTCAAGGGTCATTAACTGCAGC  
ATCATCGAATAGCCAGTGCCGTACAGTGAAAATCACTATATTTGGTGACGAAGGAGTCCCAGTGCAAGTG  
GATGGTGAAGCGTGGGTTTCCAGCTCCAGGGATTATCAAAATTGTGCACAAAAACAGAGCACAAATGCTAAC  
AAGGACAGAGCCTTTGAGAGCACTCTGAAATCTTGGGAAGATAAGCAGAAGTGTGATTCTGGTAAACCAG  
TTCTCCGAACCCATTTGTACATCCATCACGCCATTGACTTGGCAACAGAAGAGGTGTGCGAGATGCAGCTA  
TGCTCCCAGGCTGCAGAGGAGCTCATTACTAGGATATGTGACGCAGCCACAATTCACTGTCTTTTGGAGCA  
AGAACTGGCCCATGCTGTGAATGCCTGCTCCCATGCCCTGAATAAAGCCAACCAAGGTGCCCGGAGAGTC  
TTACAAGAGACACTGCCACTGAAATAGCCATCAATGTGAAGGCGCTGTATAATGAACAGAATCTTTGCTA  
GTTGGCAGGGTTTCTTTGACGCTGGAATCGCCACATGAAGAGCGAGTATCCAATGCTTACACTCTGTGGA  
GGTGAATTACAGAACTGACAGAGATTCTTGGCTTTATTATATCTTACCCCAAATGAAGATGAGGAAC  
CTCTATGGATTGCACCAAAGGAACAACAGAAGCACCGTATTTTGAATAGTGCCAAAGTTTAAAAAGGAA  
AAGGTTTCAAGAGCAGAAGACAAGTTCACAGCCTGGATCTGGGGATACCGAAAGTGGGTCATGTGAAGCGAA  
TTCTCCAGGGAATTAA

>SGK007\_ID#NA\_56

ATGGCTCTCAAACCATGTCTTGAGGCCCCAATTGAGTCTGGGCTCTGTTTCAGGGATTGAATGTGATGCTGA  
GCACCCAAACACTAGTTCTGATGCTCTTTGCCAGTGTGCTGGTGAAGTTCGCTTGGAGGCTGCTAAACTCACCG  
TAGGCTTCCAGACCCCTTGAACATCTCCCATCCATTGAGCATGCAAAGGCTGGGTGCAGGCCTCCAGATT  
GCCATGGACAAGGTCAACTCAGAGCTAGTGGACCTTGGAAATTTACCACTTACACCAACTCAGCCTGCAG  
CACCAAGGAGTCTCTTGCTATTTTCAACAACAGGTCCAGAACGAACAGATTCTGCCCCTGTTTGGACCAG  
CATGCCCAGAAGCAGCTGAGGACATTGGTGATGTGCTCCAGGAAAGTCTCCAGTACCTGGGCTGGAAACAC  
ATTGGGATGTTTGGAGGCTACTCTGGGGCTTCTCTCTGGGATGGAGTGGATGAAGTGTGGAGGGTTGTAGA  
GAATGAACCTCAAATCCCATTATCATCACTGCCAGCATGAGATACCAACAATAGTCTAGTCTCTCTTC  
AAGAGCATCTTTGGAGGATATCATCAATTGCCAGGGTTATCATCTTAACGTGCAGCTCAGAGGTCAGAAAA  
ATTATTCTTCTGGCTGCAGCAAACCTGGGACTCAACACTGGAGAATTTGTTTTCATCATTTTGCAGCAGTT  
GGAGGTGGGTACCGTTGACCACTCAATGGCTCTGGTAGTAAGAGGACCGTTTTTGAAGGAAGTACTGACA  
AATCAGAAAATGATGCACCTCCAAAAGGTGTATGGGTCACTGCTTCTCATTTGCCCTCAGCTCCTACAGAAA

Figure 1DD

AGGCCGTGGAGACGAAGGCTTTTGGAAACAAGTCTACCAAACACCGAGGAGGTGCGCCCTTCCGCAGCACCA  
TCTCCTGGGAGGAGCAGGAGGATGCCAGGACCTAGACCCAGGATGGAGTGGTCAGAGATGTGGAGCTAGGA  
CAGGTGAGGGCCAGCAGGTGAGCCCTTACTCTGCCTACCTTCACGACGCCGTCTGCTCTATGCTGAGACC  
GTGAAGCAGGTGGTAAAGGCTGGAGGCGACTTCCAGGATGGGTGGCAGCTGGTCAGCGCTCTGAAGGGTTC  
CAGTCAGACCACAGTGCAGGGAATCACAGGCCCTGTGTTTGTGGATGCCAGGGAGAAAGGCACATGGATT  
ACTCTGTCTATGCCCTGCAGAAGTCTGAAAATGGGCCCCCTTTTACTTTCTTTCTTCATTATGACAGTTAT  
CAAAGCGTGAAGTGTGCTTCTGACATTGATGATCCTCATTCTGTCTTTGGGAGCTGCCATCATAGGTCT  
GATTTTAAAGGATGCAGAGGCAAAACAAAGACATCTGGTGGCAAATCAATTTTGATGATATCACCATTCTTC  
CCCAGAACAAGCCATCCCAGAGAGCCACACCTGTGTCAAAGGCATCAACAGTAACTCATCTAGTGTGATG  
ATTTCTGTGGACCTCAGCTCTTTTGTCAAGAGCCAGCAGTGGGAAGAGCTCTTCTATGCCGCAGTAGGGCT  
TTATCAGGGAACCATGTGGCCATCCGTTACGTTGGTGACCAAGCAGAAGCCTGGGTAGGAAGCCGATTG  
TGCTACAGGAAATACAGCTGCTTGTGCTTACTTTTACCACCAGAAGCGACAAAGCCAGTGACCTGGGA  
GTGAGTCCGGGCATGCTGTTCTCCACAGGAGCCCCCTGGGCTCCACAGCAACCTGAAACCTTCCAATG  
CCTGATGGATGGTGGCTGCGCTGCAGTACCCCTCACTCCATGCGGAATCCCATCTTTTGCAGAGCTCTACTGGA  
CTGCCCCAGAGCTGCTGCAGTTCCCGGAGATGCCCTGGTTCGGGTACCCCGCAAGGAGATGTTTACAGCTTC  
GCCATTCTGATGAGGGAGCTGATCTACCATTGGGACCACGGGCCCTTTTGATGACCTCCACGAGGCACCGGA  
TGAAATCATCAACCAATCAAAGACCCCTGCAGCAGCAGTCCCACTGCAACCTTCCCTGCCCGAGGAGAAGG  
GCAATGAAAAGATCGTGGCCATGGTGAGGGTGTGTTGGGATGAATCTCTGGAGAAAAGACCCAGTTTCTCT  
TCCATCAAGAAAACCTTACGAGAGGCCAGTCCCAGAGGTGATGTGAGCATACTGGCCAGTGTGATGAGCAA  
GCTGGAAGTGTATGCCAATTACCTGGAGGAAGTGGTGCAAGAGAGGACCAGCCAGCTGACTGCAGAGAAGA  
GGAAGGTGGAGAAGCTTCTGTCCACCAAAGTGCCAGCTTCACTGGAGAACAATACTAGCCGGAAGGTCC  
GTGGAACCAGAACATTTTGAATCTGTGACAATCTTTTATCTGATATTGTTGGATTACAAAGCTGTGTTT  
TCTCAGCTCCCCCTTTGCAAGTCGTCAAGCTCCTCAATGACGTGTACAGTTTATTGATCACATCATCACAA  
CTTATGATGTTTATAAGGGCAAAGGAGAGCAAACAACCTTTCTGGTTGAAAGATAAAGAAGGCTTCACTCTT  
CCACTCCCGAATTTACTGAGGAAAAAGCCAAAGTCCCAGAGATATTGTGAGCTAATCAGCTTGCAAGGAGA  
TAGGAGCTCATATCTGACAGCATCT

>SGK050\_ID#NA\_57

AGAGAATTAGATCATCCAAACATCTGCAGATTCAGTGGAGGCTGTATTGCACTGCCAGATGTGGTTATTGT  
GATGGAATACTGCCCCAAAGGGAAGCTTGATGGATGTTCTGCTCAATGATAACATCTCTTTCAACTGGGGAT  
TT

Figure 1EE

&gt;SGK187\_CRIK\_ID#AA\_58

MLKFKYGARNPLDAGAAEPIASRASRLNLFQGGKPPFMTQQQMSPLSREGILDALFVLFECCSQPALMKIK  
HVSNFVPEVYSDTIAELQELQPSAKDFEVRSLVCGHFAEVQVVREKATGDIYAMKVMKKKALLAQEQVSF  
FEEERNILSRSTSPWIPQLQYAFQDKNHLVMEYQPGDLSLLNRYEDQLDENLIQFYLAELILAVHSV  
HLMGYVHRDIKPENILVDRTGHIKLVDFGSAAKMNSNMVNAKLPIGTPDYMAPEVLTVMNGDGKGTGYGLD  
CDWWSVGVIAIYEMIYGRSPFAEGTSARTFNNIMNFQRLKFPDDPKVSSDFDLIQSLLCGQKERLKFEGGL  
CCHPFFSKIDWNNIRNSPPFVPTLKSDDDTSNFDEPEKNSWVSSSPCQLSPSGFSGEELPFVGFYSKAL  
GILGRSESVVSGLDSPAKTSSMEKKLLIKSKELQDSQDKCHKMEQEMTRLHRRVSEVEAVLSQKEVELKAS  
ETQRSLEQDLATYITECSSLKRSLEQARMEVSDQEDDKALQLLHDIREQSRKLQEIKEQEYQAQVEEMRLM  
MNQLEEDLVSARRSDLYESELRESRLAAEEFKRKATECQHKLKAKDQKGPEVGEYAKLEKINAEQQLKI  
QELQEKLEKAVKASTEATELLQNIQAKERAERELEKLQNRDSSSEGIRKKLVEAEERRHSLENKVKRLET  
MERRENRLKDDIQTSSQIQQADKILELEEKHREAVSAQHLEVHLKQKEQHYEEKIKVLDNQIKKDLAD  
KETLENMMQRHEEEAHEKGKILSEQKAMINAMDSKIRSLEQRIVELSEANKLAANSSLTQORNMKQAQEMI  
SELRQQKFYLETQAGKLEAQRNKLQEEQLEKISHQDHSKNRLLLELETRLEVSLEHEEQKLEKROLTELQ  
LSLQERESQLTALQAARALESQRLQAKTELETTAAEIEEEIQAHTARDEIQRFDAALNSCTVITDLEE  
QLNQLTEDNAELNNQNFYLSKQLDEASGANDEIVQLRSEVDHLRREITEREMQLTSQKQTMELKTTCTML  
EEQVMDLEALNDELLEKERQWEAWRSVLGDEKSQFEQVRELQRMLEDTEKQSRARADQRTESRQVVELAV  
KEHKAELALQALKEQKLKAESLSDKLNDELEKHAMLEMNARSLOQKLETERELKQRLLEEQAQLQQQMD  
LQKNHIFRLTQGLQALDRADLLKTERSLEYQLENIQVLYSHEKVKMEGTISQOTKLIDFLQAKMDQPAK  
KKKVPLOYNELKLALEKEKARCAELEALQKTRIELRSAREEAHRKATDHPHPSTPATARQQIAMSIVR  
SPEHQPSAMSLAPPSSRRKESSTPEEFSRRLKERMHNIPHRFNVGLNMRAATKCAVCLDTHVHFGQASKC  
LECQVMCHPKCSTCLPATCGLPAEYATHFTEAFCDKMNSPGLQTKPESSSLHLEGWVKVPRNNKRGQQGW  
DRKYIVLEGSKVLIYDNEAREAGQRPVEEFELCLPDGDSIHGAVGASELANTAKADVPIYILKMESHPTT  
CWPGRTLYLLAPSPFDKQRWVTALESVVAGGRVSREKAADAKLLGNSLLKLEGDDRLDMNCTLPFSDQVV  
LVGTEEGLYALNVLKNLSLTHVPGIGAVFQIYIIKDLEKLLMIAGEERALCLVDVKKVKQSLAQSHLPAQPD  
ISPNI FEAVKGCHLFGAGKIENGLCICAAMPKVVILRYNENLSKYCIRKEIETSEPCSCIHFTNYISILIG  
TNKFYEIDMKQYTLLEFLDKNDHSLAPAVFAASSNSFPVSIVQVNSAGQREEYLLCFHEFGVFDVSYGRRS  
RTDDLKWSRLPLAFAYREPLYLVTHFNSLEVEIQARSSAGTPARAYLDIPNPRYLGPASSGAIYLAASSY  
QDKLRVICCKGNLVKESGTEHHRGPSTSRSPNKRGPPTYNEHITKRVASSPAPPEGSPHPREPSTPHRYR  
EGRTELRRDKSPGRPLEREKSPGRMLSTRRERSPGRLFEDSSRGLPAGAVRTPLSQVNKVRQHS

&gt;SGK064\_GRK7\_ID#AA\_59

MVDMGALDNLANTAYLQARKPSDCDSKELQRRRRSLALPGLQCAELRQKLSLNFHSLCEQQPIGRRLFR  
DFLATVPTFRKAATFLEDVQNWELAEEGPTKDSALQGLVATCASAPAPGNPQPFLSQAVATKQAAATTEEE  
RVAAVTLAKAEAMAFLOEQPFKDFVTSFAFYDKFLQWKL FEMQPVSDKYFTEFRVLGKGGFGEVCAVQVKN  
GKMYACKKLDKRLKKGGEKMALEKEILEKVSSPFIVSLAYAFESKTHLCLVMSLMNGGDLKFHIYNVG  
TRGLDMSRVIFYSAQIACGMLHLHELGIYVRDMKPENGLLDDLGNCRLSDGLAVEMKGGKPIQORAGTNG  
YMAPEILMEKVSYSYPVDWFAMGCSIYEMVAGRTPFKDYKEKVKSKEDLKQRTLQDEVKFQHDNFTTEAKDI  
CRLFLAKKPEQRLRSREKSDPRKHFFKTINFPRLEAGLIEPPFVDPDSVYAKDIAEIDDFSEVRGVEF  
DDKDKQFFKNFATGAVPIAWQEEIIETGLFEELNDPNRPTGCEEGNSSKSGVCLLL

&gt;SGK409\_KIAA0303\_ID#AA\_60

MQTKHFLVVTHVTPKHLHPSTCSAVTSYQCSVHACSPAPSGAASPPSSCRTSNRKSILIGNQSPALPRP  
HSPLSAHAESLFEPYMRIVCNLLTRFSYISSTPLNLATRSLEILNWTVEIALKIVQEISPPVPQPIFHL  
HGDAILHLPLGLLHPSHMSLSSSAQEKHRENKMAPSAKFLNIQDKLGLWSECPRYLFSFPQRPNYILKWKI  
EKLHQLPYQPTPDELHFLSKHFCTTESIATENRCRNTPMRPRSRSLRHCISKIPINLEEKALQSCSPVRLAT  
AQMEERLKEIITSYSPDNVPLADGVLSFTHHQIIEIARDCLDKSHQGLITSRYFLELQHKLDKLLQEAHD  
RSESGELAFIKQLVRKILIVIPARLLECLEFDPPEFYLLLEAAEGHAKGQGIKTIDIPRYIISQLGLNK  
DPLEEAMHLGNYDSGTAETPETDESVSNSASLKLRRKPRESDFETIKLISNGAYGAVYVRHKEQRQFA  
MKKINKQNLILRNQIQAFVERDILTFAENPFVVS MYCSFETRRLCMVMEYVEGGDCATLMKNMGPLPVD  
MARMYFAETVLALEYLHNYGIVHRDLKPDNLLVTSMGHIKLTDFGLSKVGLMSMTNLNIEGHIKEDAREFL  
DKQVCGTPEYIAPEVILRQGYGKPVDDWAMGIIIEFLVGCVPFFGDTPEELFGQVISDEINWPEKDEAPP  
PDAQDLITLLLRQNPLERLTGGAYEVKQHRFFRSLDWNLSLLRQKAEFIPQLESEDDTSYFDRSEKYHHM  
ETEEEDDTNDEDFNVEIRQFSSCSHRFSKVFSIIDRITQNSAEKEDSVDKTKSTLPSTETLSWSSEYSE  
MQQLSTSNSSDTESNRHLSSGLPKLAISTEGEQDEAASCPGDPHEEPGKALPPEECAQEEPEVTPAS  
TISSTLSVGSFSEHLDQINGRSECVDSTDNSSKPSSEPAHMARQRLSETEKKKISGVTKSLASALS  
MIPGDMFAVSPGLSPMSPHSLSSDPSSSRDSSPSRDSASASPHQPIVIHSSGKNYGTIRAIRVYVGDS

Figure 2A

DIYTVHHIVWNVEEGSPACQAGLKAGDLITHINGEPVHGLVHTEVIELLLKSGNKVSIITTPFENTSIKTG  
 PARRNSYKSRMVRRSKSKKKESLERRRSLFKKLAKQPSPLHTSRFSCLNRSLSGSESLPGSPHSLSP  
 RSPTPSYRSTPDFPSGTNSSQSSSPSSAPNSPAGSGHIRPSTLHGLAPKLGGQRYRSGRRKSAGNIPLSP  
 LARTPSPTPQPTSPQRSPSPLLGHSLGNSKIAQAFPSKMHSPTIVRHIVRPKSAEPPRSPLLKRVQSEEK  
 LSPSYGSDKKHLCRKHSLVLTQEEVQREQSQREAPLQSLDENVCVPLSRARVPEQGCLKRPVSRKVGR  
 QESVDDLDRDKLAKVVKADGFPEKQESHQKSHGPGSDLENFALFKLEEREKKVYPKAVERSSTFENKA  
 SMQEAPPLGSLKDALHKQASVRASEGAMSDGPVPAEHRQGGGDFRRAPAGTLQDGLCHSLDRGISGKGE  
 GTEKSSQAKELLRCEKLDKLANIDYLRKKMSLEDKEDNLCVPLKPKMTAGSHECLPGNPVRPTGGQOEPP  
 PASESRAVFSSTHAAQMSAVSFVPLKALTGRVDSGTEKPGLVAPESPVRKSPSEYKLEGRSVSCLEPIEGT  
 LDIALLSGPQASKTELPSPEAQSPSPSGDVRAVPPVLPSSSGKNDTTSARELSPSSLKMNKSYLLEPW  
 FLPPSRGLQNSPAVSLPDPEFKRDRKGPHPTARSPGTVMESNPQQREGSSPKHQDHTTDPKLLTCLGQNLH  
 SPDLARPRCLPPEASPSREKPGRESSERGPPTARSERSAARADTCREPSMELCFPETAKTSDNSKNLLS  
 VGRTHPDFYTQTQAMEKAWAPGGKTNHKDGPGEARPPPRDNSSLHSAGIPCEKELGKVRRGVEPKPEALLA  
 RRSLLQPPGIESEKSEKLSFSLQKDGAKEPERKEQPLQRHPSSI PPPPLTAKDLSSPAARQHCSSPSHAS  
 GREPGAKPSTAEPSSSPQDPPKPVAAHSESSSHKPRPGPDGPPKTKHPDRSLSSQKPSVGATKGKEPATQ  
 SLGGSSREGKGHSKSGPDVFPATPGSQNKASDGIGQGGGSPVPLHTDRAPLDAKPQPTSGGRPLEVLEKP  
 VHLPRPGHPGPSEPADQKLSAVGEKQTLSPKHKPKSTVKDCPTLCKQTDNRQTDKSPSQPAANTDRRAEGK  
 KCTEALYAPAEQKLEAGLSFVHSENRLKGAERPAAGVGKGFPEARAGKPGPGKPPTEADKPNGMKRSPSA  
 TGQSSFRSTALPEKSLSCSSSPETRAGVREASAASSDTSSAKAAGGMLELPAPSNRDRHKAQPAGEGRTH  
 MTKSDSLPSFRVSTLPLESHHPDNTMGGASHDRALSVTATVGETKGKDPAPAQPPPPARKQNVGRDVTKP  
 SPAPNTDRPISLSNEKDFVVRQRRGKESLRSSPHKKAL

>SGK021\_ID#AA\_61

MGANTSRKPPVFDENEDVNFDFEILRAIGKGSFGKVCIVQKNDTKKMYAMKYMKNQKQKVERNEVRNVFKE  
 LQIMQGLEHPFLVNLWYSFQDEEDMFVVDLLLGGDLRYHLQQNVHFKEETVKLFICELVMALDYLQNQRI  
 IHRDMKPDNILLDEHGHVHITDFNIAAMLPRETQITTMAGTKPYMAPEMFSSRKAGYSFAVDWWSLGVT  
 YELLRGRPHYHRSSTSSKEIVHTFETTVVTYPSAWSQEMVSLKKVRRKTACPNEVTKGSRLSGLSLEVS  
 IQYWGHLQSSLOKGRNLCDPTFELEEMILES KPLHKKKKRLAK

>SGK410\_ID#AA\_62

MAAVLRQARAAGSPTASEAPSPHPRPPALRRGSEEMPTQRDSSTMSHRVAGGCSGDHSHQVWVKAYYRGDI  
 MITHFEPVSFEGLYNVVRDMCSFDNEQLFTMKWIDEEGDPCTVSSQLELEEAFTLYELNKDSELLIHVFP  
 CVPERPGMPCPREDKSIYRRGARHWRKLYCANGHTFQAKCFNRRADCAICTDRIWGLGCQGYKCINCKLLV  
 HKKCHKLLTIECGRHSLPLEPMPMDQSSMHSMDHAQTVPYNPSSSHESLDQVGEKEAMNTRRESGKASSSL  
 GLQDFDLLQVIGRGSYAKVLLVRLKKTORIYAMKVVKELVNDDEDIDWVQTEKHVFEQASNHFFVLGLQS  
 CFQTESRLLFFVIEYVNGGDLFMHMQRQKLPPEEHARFYSAEISLAFNYLHERGIYRDCLKDNVLLDSEGH  
 IKLTDYGMCKEGLRPGDITSTFCGTPNYIAPEILRGGDYGFSDWWALGVLMEFMETMAGRTPFDTVGSNDP  
 DQNTEDSLFQVIVEKQIHIPRSLSGKAAASVLKSLFNKDPKKRLGCHPQTGFADIQGHFFQNVWDMMMEQK  
 QVVSFPKPSISGEFGLDNFDSQFTNEPVWLTDPVDDIVRKIDQSEFEGFEYINPLMSAEECV

>SGK069\_ID#AA\_63

PSQFCSHRAPRIVIVPAPSPFLRPVNPKSQAWYPWVSFLTKASPRPAVPWHLPDRTSLACGGSRLQDGD  
 LDPNPRPQHPAEGPELETVPFKQEEMPGKQSEEGPAEAGASEDSEEEGLGGLTLEELQQGQEAARALEDMM  
 TLSAQTLLVRAEVDLYEEVRPLGQGRYGRVLLVTHRQKGTPLALKQLPKPRTSLRGFLYEFVGLSLGAHS  
 AIVTAYGIGIESAHSYSFLTEPVLHGDLMFAIQPKVGLPQPAVHRCQAQLASALEYIHARGLVYRDLKPN  
 VLVCDPACRRFKLTDFGHTRPRTLLRLAGPPIPYTAPELCAPPPLPEGLPIQPALDAWALGVLLFCLLTG  
 YFPWDRPLAEADPFYEDFLIWQASGQPRDRPQWFGGLAPAADALLRGLLDPHPRRRSAVIAIREHLGRPW  
 QREGAEAVGAVEEEAGQAPAVATEVLPAGAGLTAGVLLWSDPKPSNPGSSSHRVGASDLNPGMSCRAQRE  
 EGLGGTGVWEQTVHLSGPPDVIRTPPPPRPSNWTLLAFAGLRGRQPSSTTSKPREEEEEEMVSCRLRLQD  
 PESE

>SGK110\_ID#AA\_64

MHPHSGLGAPGLLPQTGAGGASVAVTPNLSRTQKQVARVREDTATALQRLVELTTSRVTPVRSLRDQYHLI  
 RKLGSYSYGRVLLAQPHQGGPAVALKLLRRDLVLRSTFLREFCVGRCSAHPGLLQTLAGPLQTPRYFAFA  
 QEYAPCGDLSGMLQERGLPELLVKRVVAQLAGALDFLHSGLVHADVKPDNVLVFDPVCSRVALGDLGLTR  
 PEGSPTPAPPVPLPTAPPELCLLLPPDTLPLRPAVDSWGLGVLLFCAATACFPWDVALAPNPEFEAFAGWV

Figure 2B

TTKPQPPQPPPPWDQFAPPALALLQGLLDLDPETRSPPLAVLDFLGDDWGLQGNREGPGVLGSAVSIEDRE  
EGSSLEEWTDEGDDSKS

>SGK053\_CKLIK\_ID#AA\_65

MARENGESSSSWKKQAEIDIKKIFEFKETLGTGAFSEVVLAEKATGKLFVAVKCIKKALKGKESSIENEIA  
VLRKIKHENIVALEDIYESPNHLYLVMQLVSGGELFDRIVEKGFYTEKDASTLIRQVLDVAVYYLHRMGIVH  
RDLKPENLLYYSQDEESKIMISDFGLSKMEGKGDVMSTACGTPGYVAPEVLAQKPYSAVDCWSIGVIAYI  
LLCGYPFFYDENDSKLFEQILKAIEYFDSPLYWDDISDSAKDFIRNLMEKDPNKRYTCEQAAHPWIAGDTA  
LNKNIHESVSAQIRKNFAKSKWRQAFNATAVVRHMRKLHLGSSLDSSNASVSSSLASQKDCAYVAKPES  
LS

>SGK124\_ID#AA\_66

MRATPLAAPAGSLSRKKRLELDNLDTERPVQKRARGPQPRLPCLLPLSPPTAPDRATAVATASRLGPY  
VLEPEEGGRAYQALHCPGTGTEYTKVYPVQEAALVLEPYARLPPHKHVARPTEVLQAGTQLLYAFFTRTHG  
DMHSLVRSRHRIPPEAAVLFQMATALAHCHQHGLVLRDLKLCRFVFAERERKKLVLENLEDSCLVTGPD  
DSLWDKHACPAYVGPEILSSRASYSKGAADVWSLGVALTMLAGHYPFQDSEPVLLFGKIRRGAYALPAGL  
SAPARCLVRCLLRREPAERLTATGILLHPWLRQDPMPLAPTRSHLWEAAQVVPDGLGLDEAREEEGDREV  
LYG

>SGK254\_CAMKKA\_ID#AA\_67

MEGGPAVCCQDPRAELVERVAIDVTHLEADGGPEPTRNGVDPPPRARAASVPGSTSRLPARPSLSAR  
KLSLQERPAGSYLEAQAGPYATGPASHISPRARRPTIESHHVAISDAEDCVQLNQYKLQSEIGKVGLTDA  
YLQAGYGVVRLAYNESEDRHYAMKVLKSKLLKQYGFPRPPRGSAQAGGPAKQLLPLERVYQETAILK  
KLDHVNVLKIEVLDDPAEDNLYLVDLLRKGPVMEVPCDKPFSEEQARLYLRDVILGLEYLHCQKIVHRDI  
KPSNLLGDDGHVKIADFGVSNQFEGNDAQLSSTAGTAPFMAPEAISDSGQSFSGKALDVWATGVTLYCFV  
YGKCPFIDDFILALHRKIKNEPVVFPPEPEISEELKDLILKMLDKNPETRIGVPDIKLHPVWTKNGEPLP  
SEEHCSVVEVTEEEVKNSVRLIPSWTTVILVKSMLRKRSFGNPFEPQARREERSMSAPGNLLVKEGFGE  
GKSPELPGVQEDEAAS

>SGK297\_CAMKIB2\_ID#AA\_68

MLLLKKHTEDISSVYERIRERLGGAFSEVVLQERGAHLVALKCIKKALRGKEALVENEIAVLRIRISHP  
NIVALEDVHESPSHLYLAMELVTTGGELFDRIMERGSYTEKDASHLVGQVLGAVSYLHSLGIVHRDLKPENL  
LYATPFEDSKIMVSDFGLSKIQAGNMLGTACGTPGYVAPELLEQKPYKAVDVWALGVISYILLCGYPFFY  
DESDPELFSQILRASIEYFDSFPGWDISESAKDFIRHLLERDPQKRFTCQALRHLWVSGDTAFDRDILGSV  
SEQIRKNFARTHWKRAFNATSFLRHIRKLQIQPEGEGASEQGMARHSHSGLRAGQPPKW

>SGK411\_CAMKII\_DELTA2\_ID#AA\_69

MASTTTCTRTDEYQLFEELGKGAFSVVRRCKIPTGQGYAAKIINTKKLSARDHQKLEREARICRLKHP  
NIVRLHDSISEEGFHYLVFDLVTGGELFEDIVAREYYSEADASHCIQQILESUNHCHLNGIVHRDLKPENL  
LLASKSKGAAVKLADFGLAIEVQGDQAWFGFAGTPGYLSPEVLRKDPYKGPVDMWACGVILYILLVGYPP  
FWDEDQHRLYQKIKAGAYDFPSPWDVTTPKADLINKMLTINPAKRITASEALKHPWICQSTVASMMHR  
QETVDCCLKFNARRKLKGAILTTLATRNFSAAKSLKKPDGVKESTESSNTTIEDVDKARKQEIIVKTE  
QLIEAINNGDFEAYTKICDPGLTAFEPEALGNLVEGMDFHRFYFENALSKSNKPIHTIILNPHVHLVGDDA  
ACIAYIRLTQYMDGSGMPKTMQSEETRVWHRRDGKWQNVHFHRSQSPTVPIKPPCI PNGKENFSGGTSLWQ  
NI

>SGK027\_ID#AA\_70

MTAVYMNGGGLVNPYARWDRRDSVESGCQTESSKEGEGQPRQLTPFEKLTQDMSQDEKVVREITLGKRI  
GFYRIRGEIGSGNFSQVKGILHSLTKEKVAIKILDKTCLDQKTORLLSREISSMEKLHHPNIIRLYEVVET  
LSKLHLMVEYAGGGELFGKISTEGKLESEPESKLIFSQIVSAVKHMHENQIIHRDLKAENVFYTSNTCVKVG  
DFGFSTVSKKGEMLNTFCGSPPYAAPELFRDEHYIGIYVDI WALGVLLYFMVTGTMFPRAETVAKLKKSIL  
EGTYSVPPHVSEPCRLIRGVLQQIPTERYGIDCIMNDEWMQGVYPTPLEPFQLDPKHLSETSTLKEEEN  
EVKSTLEHLGITEHIRNNQGRDARSSITGVYRIILHRVQRKKALESVPVMMPLDPKERDLKKGSRYRGI  
RHTSKFCSIL

>SGK046B\_ID#AA\_71

Figure 2C

EEEEAQTMTFGQILLAMXYCLYKNVAHADVSLQIILLPEERNIKIVDFGFSMTFGERRFGNPFCEGMPYVAT  
ELFL

>SGK046C\_ID#AA\_72  
FRQILSGLQYCRLKNITHPELKPQTHLVIEEGNIKITDF

>SGK089\_ID#AA\_73  
NIFFDKERNIKIANYGFSSTFAKRDNTEKGTFDGTLPYAGPELFLGHGYQCPAINVWSLGIIVCKMVAGAL  
PFYEQNCKDILKK

>SGK133\_ID#AA\_74  
MVAGLTLGKGPEPDGDVSVPERKDEVAGGGGEEEEAEERGRHAQYVGPYRLEKTLGKGQTLVKGVLGHCI  
TGQKVAIKIVNREKLSVLMKVEREIAILKLIHHPVLKLDVYENKKYLYLVLEHVSGGELFDYLVKKG  
RLTPKEARKFFRQIVSALDFCHSYSICHRDLKPENLLDEKNNIRIADFGMASLQVGSLLTSCGSPHYA  
CPEVIKGEKYDGRADMWSCGVILFALLVGALPFDDNLRQLLEKVKRGVFMHPHFIPDQCSSLRGMIEV  
EPEKRLSLEQIQKHPWYLGKGHEPDPCLEPAGRRVAMRSLPSNGELDPDVLESMAISGCFRDRERLHREL  
RSEENQEKMIYLLDRKERYPSCEDQDLPPRNDVDPKRKRVDSPLSRHGKRRPERKSMEVLSITDAGG  
GGSPVPTRRALEMAQHSQRSRSVSGASTGLSSSPSPVFSFSPEPGAGDEARGGGSPTSSTQTLPSR  
GPRGGGAGEQPPPSARSTPLPGPPGSPRSSGGTPLHSPHPTPRASPTGTPTGTPPPSPGGVGGAAWRSR  
LNSIRNSFLGSPRFHRRKMVPTAEEMSSLTPESSPELAKRSWFGNFISLDKEEQIFLVLKDKPLSSIKAD  
IVHAFLSIPSLSHSVLSQTSFRAEYKASGGPSVFQKPVRFQVDISSSEGPEPSRRDGGSGGGIYSVTFTL  
ISGSPRRFRKRVETIQALLSTHDQPSVQALADEKNGAQTRPAGAPPRSLQPPPGRPDPSELSSSPRRGPPK  
DKKLLATNGTLP

>SGK004\_MSK\_ID#AA\_75  
MVIMSEFSADPAGQGQGGQKPLRVGFYDIERTLGKGNFAVVKLARHRVTKTQVAIKIIDKTRLDSSNLEKI  
YREVQLMKLLNHPHIIKLYQVMEKDMLYIVTEFAKNGEMFDYLTSGHLSSENEARKKFWQILSAVEYCHD  
HHIVHRDLKTENLLLDGNMDIKLAGTEDFGFGNFYKSGEPLSTWCGSPPYAAPEVFEGKEYEGQLDIWSL  
GVVLYVLVCGSLPFDGPNLPTLRQRVLEGRFRIFFMSQDCESLIRMLVVDPAARRITIAQIRQHRWMRAE  
PCLPGPACPAFSAHSYTSNLGDYDEQALGIMQTLGVDRQRTVESLQNSSYNHFAAIYLLERLKEYRNAQ  
CARPGPARQPRRSSDLGLEVPQEGSLDTPFRPALLCPQFQTLVQSVLQAEKDCELQSSLOWPLFFPVDA  
SCSGVFRPRPVSPSSLLDTAISEEARQGGGLEEQDTQESLPSSTGRRHTLAEVSTRLSPLTAPCIVVSPS  
TTASPAEGTSSDCLTFSASKSPAGLSGTPATQGLLACSPVRLASPFGLGSQSATPVLQAQGGGLGGAVLLP  
VSFQEGRRASDTSLTQGLKAFRQQLRKTTRTKGFLGLNKIKGLARQVCQAPASRASRGGLSPFHAPAQSPG  
LHGGAAGSREGWSLLEEVLEQQRLQLQHHPAAAPGCSQAPQAPAPFVIAPCDGPGAAPLPSTLLTSGLP  
LLPPPLLQTGASPVASAAQLLDTHLHIGTGPTALPAVPPPRLARLAPGCEPLGLLQGDCEMEDLMPCSLGT  
FVLVQ

>SGK006\_ID#AA\_76  
VPLYFRKLSHDLQMTFPYTKKVAYLVGKMINKGPFKAVIEGLGLPMGEKATIKGISKKKAKQDSYVLKNI  
KCESWIQ\*IVKYPNASQLQTLTKNSYYVMELCLSGNLMNRIWDQKKLSEREVKMIFRQMLSAMEHVHWH  
GMVHKDYAFFFFSGRKSHQSPQVYETLDINTEKLEGLSQEQSLTQGGSPAYAAPELLAHQKYGPKVDVWFM  
GKFLMLTGTLPTLEIFISIEQLYQKMVIGAISNVLPESFGLLWSLSLV

>SGK180\_SNRK\_ID#AA\_77  
MAGFKRGYDGKIAGLYDLDKTLGRGHFACGTCSRHVFTGEKVAVKVIDKTKLDTLATGHLFQEVRCMKLVQ  
HPNIVRLYEVIDTQTKLYLILELGDGDMFDYIMKHEEGLNEDLAKKYFAQIVHAISYCHKLHVVRDLKP  
ENVVFFEKQGLVKLTDFGFSNKFQPGKLTTSGLAYSAPAILLGDEYDAPAVDIWSLGVILFMLVCGQP  
PFQEANDSETLTMIMCKYTVPSHVSKECKDLITRMLQRPKRASLEEIENHPWLQGVDPSPATKYNIP  
VSYKNLSEEEHNSIIQRMVLGDIADRAIVEALETNRYNHITATYFLAERILREKQEKIOTRSASPSNI  
KAQFRQSWPTKIDVPQDLEDDLTATPLSHATVPQSPARAADSVLNGHRSKGLCDSAKKDDLPELAGPALST  
VPPASLKPTASGRKCLFRVEEDEDEEDKPKMSLSTQVVLRRKPSVTIRLTSRKSAFVLNQIFEEGESDD  
EFMDENLPKLSRLKMNIASPGTVHKRYHRRKSQGRGSSCSETSDDDESERRLDKDSGFTYSWHRRD  
SSEGPPGSEGDGGGQSKPSNASGGVDKASPSENNAGGSPSSGSGGNPTNTSGTTRRCAGPSNSMQLASRS  
AGELVESLKLMSLCLGSQHLHGSTKYIIDPQNGLSFSSVKVQEKSTWKMCSSTGNAGQVPAVGGIKFFSDH  
MADTTTELERIKSKNLKNNVLQLPLCEKTISVNIQRNPKEGLLCASSPASCCHVI

Figure 2D



>SGK386\_MLCKS\_ID#AA\_78  
MATENGAVELGIQNPSTGKERRVDPDGCLTSSADKAPKGPPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAK  
APASEKGDGTLAQPSTSSQGPKEGDRGGGPAEGSAGPPAALPQQTATPETS VKKPKAEQASGSQDPGPKP  
RVGKKAEGQAAARRGSPAFLHSPSCPAIISSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEKG  
NILAESQKEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDCCPPPPAPFP  
HRMVELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLEIEVMNQLNH  
RNLIQLYAAIETPHEIVLFMEYIEGGELFERIVDEDEDYHLTEVDTMVFVRQICDGLFMHKMRVLHLDLKE  
NILCVNTTGHVVKIIDFGLARRYNPNEKLKVNFGTPEFLSPEV VNYDQISDKTDMWSMGVITYMLLSGLSP  
FLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKQDQARMNAAQCLAHPLNLAEKAKRCNR  
RLKSQILLKKYLMKRRWKNFIAVSAANRFKKISSSGALMALGV

>SGK003\_ID#AA\_79  
MTNNSGSKAELVVGGKYKLVKRKIGSGSFGDVYLGITTTNGEDVAVKLESQKV KHPQLLYESKLYTILQGGV  
GIPHHWYGQEKDNNVLMDDLGP SLEDLFNFCSSRFTMKT VMLADQMISRIEYVHTKNFLHRDIKPDNF  
LMGTGRHCNKLFLIDFGLAKKYRDNRTQHIPYREDKHLIGTVRYASINAHLGIEQSRDDMESLG YVFM  
FNRTSLPWQGLRAMTKKQYKIESEKKMSTPVEVLCKGFPAEFAMYLNYCRGLRFEVVPDYM YLRQLFRIL  
FRTLNHQYDYTFDWTMLKQKAAQQAASSSGQGGQQAQTQTGKQTEKNKNNVKDN

>SGK066\_ID#AA\_80  
MSGGGEQPDILSVGILVKERWKVLRKIGGGGGEIYDALDMLTRENVALKVESAQQPKQVLKMEVAVLKKL  
QGGFVQIRIYIQGRHIALYFVAMSQFGRNLADLRSSQSRGTFTISTTLRLGRQILESIESIHSVGLHRDI  
KPSNFAMGRFPSTCRKCYMLDFGLARQFTNSCGDVRPPRAVAGFRGT VRYASINAHNRNEMGRHDDLWSLF  
YMLVEFVVGQLPWRKIKDKEQVGSIKERYDHRMLKHLPPFEFSIFLDHISSLDYFTKPDYQLLTSVFDNSI  
KTFGVIESDPFDWEKTGNDGSLTTTTTSTTPQLHTRLT PAAIGIANATPIPGDLLRENTDEVFPDEQLSDG  
ENGIPVGVS PDKLPGLGHPRPQEKDVWEEMDANKNIKLGICKV

>SGK041\_NKIAMRE\_ID#AA\_81  
MEMYETLGKVGESYGTVMKCKHKNTGQIVAIKIFYERPEQSVNKIAMREIKFLKQFHHENLVNLIEVFRQ  
KKKIHLVFEFIDHTVLDLQHYCHGLESKRRLKYLQILRAIDYLSNNIHRDIKPENILVSQSGITKLC  
DFGFARTLAAPGDIYTDYVATRWYRAPELVLKDTSYGKPVDIWALGCMIIEMATGNPYLPSSSDLDLHKKI  
VLKVGNLSPHLQNI FSKSPIFAGVVL PQVQHPKNARKKYPKLNGLLADIVHACLQIDPADRISSD LHHHE  
YFTRDGFIEKFMP ELKAKLLQEAKVNSLIKPKESSKENELRKDERKTVYTNLLSSSVLGKEIEKEKKPKE  
IKVRVIVKVGGRGDISPKKKEYEGGLGQDANENVHMPSPDTKLV TIEPPNPINPSTNCNGLKENPHCGG  
SVTMPPINLTNSNLMAANLSSNLFHPSVRLTERAKKRTSSQSIGQVMPNSRQEDPGPIQSQMEKGFNER  
TGHSDQMANENKRKLNFSRSDRKEFHFPPELPTIQSKDTKGMEVKQIKMLKRESKKTESKIPTLLNVDQN  
QEKQEGGDGHCEGKNLKRNRFF

>SGK112\_ID#AA\_82  
MEKYEKLAKTGESYGVVFKCRNKTSGQVVAVKKFVESEDDPVVKIALREIRMLKQLKHPNLVNLIEVFR  
RKRKMHVFEYCDHTLLNELERNPNGVADGVIKSVLWQTLQALNFCHIHNCIHRDIKPENILITKQGIKI  
CDFGFAQILSN SHVGRIDL PDLIDAFVPGDAYTDYVATRWYRAPELLVGD TQYGSVDIWAIGCVFAELL  
TGQPLWPGKSDVDQLYLIIRTLGKLIPRHQSIFKSN GFFHGISIPEPEDMETLEEFSDVHPVALNFMKGC  
LKMNPDRLTCSQLLESSYFDSFQEAQIKRKARNEGRNRRRQONQLLPLIPGSHISPTPDGRKQVLQLKFD  
HLPNI

>SGK038\_ERK7\_ID#AA\_83  
MCTVVDPRIVRRYLLRRQLGQAGYIVWKA VDRRTGEVVAIKKIFDAFRDKTDAQDMGFL LAPPTHPTVFL  
SLQRTFREITLLQEFGDHPNII SLLDVIRAENDRDIYLVFEFMDTDLNAVIRKGGLLQDVHVSIFYQLLR  
ATRFHSGHVVRDQKPSNVLLDANCTVKLCDFGLARSLGDLPEGPEDQAVTEYVATRWYRAPEVLLSSHR  
YTLGVDMWSLGCILGEMLRGRPLFP GTSTLHQLELILETI PPPSEEDLLALGSGCRASVLHQLGSRPRQTL  
DALLPPDTSPEALDLLRRLV FAPDKRLSATQALQHPYVQRFHCPSEWAREADVPRPRAHEGVQLSVPEYR  
SRVYQMILECGGSSGTSREKGEVSPSQ AHLHKPRADQLPSRTPVQGP RP RPQSSPGHDP AEHAKNVPR  
QNSAPLLQTALLNGERPPGAKEAPPLTSLVKPSGRGAAPSLTSQAAAQVANQALIRGDWNRRGGGVRVAS  
VQVQPPRLPPEARPGRRMFSTSALQGAQGGARALLGGYSQAYGTVCHSALGHLPLLEGHHV

>SGK158\_ID#AA\_84

Figure 2E



MAAILGDTIMVAKGLVKLTQAAVETHLQHLGIGGELIMAARALQSTAVEQIGMFLGKVQGDKEEYFAEN  
 FGGPEGEFHFVSPAAGASTDFSSASAPDQSAPPSSLGHAHSEGPAPAYVASGPFREAGFPQASSPLGRAN  
 GRLFANPRDSFSAMGFQRRFFHQDQSPVGGTLAEDIEKARQAKARPENKQHKQTLSEHARERKVPVTRIGR  
 LANFGLAVGLGFGALAEVAKKSLRSEDPGKKAVLGSSPFLSEANAERIVRTLCKVRGAALKLQMLSIQ  
 DDAFINPHPASALLTPARQPVRRQLPRLGLTWRVHCCPGLAVGLGFGALAEVAKKSLRSEDPGKKAVLGSS  
 SPFLSEANAERIVRTLCKVRGAALKLQMLSIQDDAFINPHLAKIFERVRQSADEMFPLKQMMKTLNNDLGP  
 NWRDKLEYFEERPFAAASIGQVHLARMKGGREVAMKIQYPGVAQSINSDVNNLMAVLNMSNMLPEGLFPEH  
 LIDVLRLEALCEDYQREAAACARKFRDLLKGHPFFYVPEIVDELCSPHVLTTELVSQGFPLDQAEGLSQEIR  
 NEICYNILVCLRELFEHFHMQTDPNWSNFFYDPPQHKVALLDFGATREYDRSFTDLYIQIIRAAADRRE  
 TVRAKSIEMKFLTGYEVKVMEDAHLDAILILGEAFASDEPFDFGTQSTTEKIHNLI PVMLRHRVLVPPPEET  
 YSLHRKMGGSFLLCSKLKARFPCKAMFEEAYSNYCKRQAQQ

>SGK429\_ID#AA\_85

MVAPWRVSVRVCLSHLRFCFELRQGLSLLRPSECPRDARLCWLLGLTLPKVVS LCGDVGEGAPDVL SRRRVR  
 CSGAAGAGPAESLPRAGPLGGVFLHLRLWLRAGALLVKFFPLLLLYPLTYLAPS VSTLWLHLLLKATETSG  
 PTYIKLGQWASTRRDLFSEAFCAQFSKLHVRVTPHPWTHTERFLRQAFGDDWGSILSFENREPVGSGCVAQ  
 VYKAYANTAFLETDSVQRLGRASCLPPFSHTGAVGGLRELFGYLGNGRKPPENLADQSFLERLLL PKADLV  
 GSNAGVSRAQVPGHQPEATNLISVAVKVLHPGLLAQVHMDLLMKIGSRVLGVLPGIKWLSLPEIVEEFEK  
 LMVQQIDRLRYEAQNLHVFQVNFVRNVKAVKFTPLRPFVTVREVLVETEEVSVVSSYQQAGIPVDLKRKIAR  
 LGINMLLKMIFVONFVHADLHPGNILVQGANGLSSSQEAQLQADICDTLVVAVPSSSLCPLRLVLLDAGIV  
 AELQAPDLRNFRAVFMVVMGQQRVAELILHARASECRDVEGFKTEMAMLVTOARKNTITLEKLHVSSL  
 LSSVFKLLMTHKVKLESNFASIVFAIMVLEGLGRSLDPKLDILEAARPFLLTGFPVCP

>SGK152\_SUDD\_ID#AA\_86

MDLVGVA SPEPGTAAAWGPSKCPWAI PONTISCSLADVMSEQLAKELQLEEEAAVFPEVAVAEGPFITGEN  
 IDTSSDLMLAQLQMEYDREYDAQLRREEKKFNGDSKVSISFENYRKVHPYEDSDSSEDEVWDQDTRDDPY  
 RPAKVPPTPKKGFIGKGDITTKHDEVVCGRKNTARMENFAPEFQVGDGIGMDLKLNSHVFNAKQHAYSE  
 ERRSARLHEKKEHSTA EKAVDPKTRLLMYKVMNSGMLETITGCI STGKESVVFHAYGGSMEDEKESKVIP  
 TECAIKVFKTTLNEFKNRDKYIKDDFRFKDRFSKLNPRKIHRMWAEKEMHNLARMQORAGIPCPTVLLKKH  
 ILVMSFIGHDQVPAPKLKEVKLNSEEMKEAYYQTLHLMRQLYHECTLVHADLSEYNMLWHAGKVWLIDVSQ  
 SVEPTPHGLEFLFRDCRNVSQFFQKGVKEALSERELFNAVSGLNITADNEADFLAEIEALEKMNEDHVQ  
 KNGRKAASFLKDDGDPPLLYDE

>SGK077\_ID#AA\_87

MAASLPGPGSRLFRTYGAADGRRQRRPGREAAQWFPQDRRRFFNSSGSSDASIGDPSQSDPDDPDDPDF  
 PGSPVRRRRRCPGGRVPKDRPSLTVT PKRWKLARPSLTVT PRRLGLRARPPQKCSSTPCGPLRLPPFPSRD  
 SGRLSPDLSCVCGQPRDDELGISASLFSSLASPCPGSPTPRDSVISIGTSACLVAASAVPSDLHLPEVSLD  
 RASLPCSQEEATGGAKDTRMVHQTRASLSVLFGLMNSGTPEDSEFRADGKNMRESCCKRKL VVGNGPEGP  
 GLSSTGKRRATQDSCQERGLQEA VRREHQAESVPKGRIVPRGTDRLERTRSSRKSKHQEATETSLHSHR  
 FKKGQKLKGD SFPTQDLTPLQACFWTKTRASF SFHKKKIVTDVSEVCSYTTATSLSGSLLSECSNRVPM  
 NRTSGAPSSWHSSSMYLLSPLNTLSISNKKASDAEKVYGECSQKGPVPFSSHCLPTEKLQRCCKIGEGVFGE  
 VFQTIADHTPVAIKIIAIEGPDLVNGSHQKT FEEILPEII ISKELSLLSGEVCNRTGFIGLNSVHCVQGS  
 YPPLLLKAWDHYNSTKGSANDRPDFFKDDQLFIVLEFEFGGIDLEQMRKLSLATAKSLHQLTASLAVA  
 EASLRFHRDLHWGNVLLKKTSLKKLHYTLNGKSSITPSCGLQVSIIDYTLRSLERDGI VVFCDVSMDEDL  
 FTGDGDYQFDIYRLMKKENNRWGEYHPYSNVLWLHYLTDKMLKQMTFKTCNTPAMKQIKRKIQEFHRTM  
 LNFSSATDLLCQHSFLK

>SGK093\_WNK3\_ID#AA\_88

MSQTEADLALRPPPLGTAGQPRLGPPRRARRFSGKAEPRPRSSRLSRRSSVDLGLLSWSLPASPAPDP  
 PDPPDSAGPGPARSPPPSSKEPPEGTWTEGAPVKAEDSARPELPDSAVGPGSREPLRVPEAVALERRREQ  
 EEKEDMETQAVATSPDGRYLKFDIEIGRSFKTVYRGLD TDTTVEVAWCELQTRKLSRAERQRFSEEVEML  
 KGLQHPNIVRFYDSWKS VLRGQVCIVLTELMTSGTLKTYLRRRFREMKPRVLQRWSRQILRGLHFLHSRVP  
 PILHRDLKCDNVFITGPTGSVKIGDLGLATLKRASFVIGTPEFMAPEMYEEKYDEADVYAFGMCMLE  
 MATSEYPYSECQNAAQIYRKVTSGRKPSFHKVKIPEVKEIIEGCI RTDKNERFTIQDLLAHAFREERG  
 HVELAEEDDGEKPGKLKWLRMEDARRGGRPRDNQAI EFLFQLGRDAEEVAQEMVALGLVCEADYQPVARA  
 VRERVAAIQRKREKLKARELEALPPEPGPPPATVPMAPGPPSVFPPEPEEPEADQHQPFLFRHASYSSTT  
 SDCETDGYLSSSGFLDASDPALQPPGVPSSLAESHLCLPSAFALSI PRSGPGSDFSPGDSYASDAASGLS

Figure 2F

DVGEGMGQMRPPGRNLRPRSRRLRVTSVSDQNDRVVEQCLOTHNSKMVTFRFDLDGDSPEEIAAAMVYN  
 EFILPSERDGLRRIREIIQORVETLLKRDTPGMEAAEDTLSPQEEPAPLPALPVPLPDPSNEELQSSTSLE  
 HRSWTAFTSTSSSPGTPLSPGNPFSPGTPIPSGPIFPITSPCHSPSPSPSPISQVSSNPSPHPTSSPLP  
 FSSSTPEFPVPLSQCPWSSLPTTSPPTFSPTCSQVTLSSPFFPPCPSTSSFPSTTAAPLLSLASAFSLAVM  
 TVAQSLLSPSPGLLSQSPPAPPSPLPSLPLPPPVPAGGQESPSPHTAEEVESEASPPPARPLPGEARLAPIS  
 EEGKPQLVGRFQVTSSKEPAEPLPLQPTSPTLSSGSPKPTPQLTSESSDTEDSAGGGPETREALAESDRAA  
 EGLGAGVEEEGDDGKEPQVGGSPQPLSHSPVWMNYSYSSLCLSSESESSGEDEEFWAELOSLRQKHLSE  
 VETLQTLQKKEIEDLYSRLGKQPPPGIVAPAAMLSSRRRLSKGSFPTSRNSLQRSEPPGPGIMRRNSLS  
 GSSTGSQEQRASKGVTFAGDVGRMFGVATETIEDALLHLAQONEQAVREASGRGRFREPQIVEFVFLLS  
 EQWCLEKSVSYQAVEILERFMVKQAENICRQATIQPRDNKRESQNRWALKQQLVNKFTLRLVSCVQLASKL  
 SFRNKIISNIPVLNFLQALGYLHTKEELLESELQDLVLSNFRINLPTPLAYVETLLEVLYGNGCLVPAMRL  
 HATCLTLLDLVYLLHEPIYESLLRASIENTSPSQLQGEKFTSVKEDFMLLAVGIIAASAFIQNHECWSQVC  
 TTEQDQHERVKVDEHALFVISA

>SGK074\_ID#AA\_89

SETDCYDIIEVLGKGTFGEVAKGWRRSTGEMVAIKILKNDAYRNRIKKNELKLLHCMRGLDPEEAHVIRFL  
 EFFHDAKSYLVFELLEQNLFQKENNFAPLPARHIRTVTLOVLTALARKELAIHADLKPENIMLVLDQ  
 TRCPRVKVIDFGSASIFSEVRYVKEPYIQSRFYRAPEILLGLPFCEKVDVWSLGCVMALHLGWPLYPGN  
 NEYDQVRYICETQGLPKPHLLHAACKAHFFKRNPHPDANPWQLKSSADYLAETKVRPLERRKYMLKSLD  
 QIETVNGGSVASRLTFPDREALAEHADLKSMVELIKRMLTWESHERISPSAALRHPFVSMQQLRSAHETTH

>SGK087\_ID#AA\_90

MNTPLQSAHHTLTLVEFIAYLVSDSPFPAASTLSLPTCNFDCPTKLGAETMPAAAGQRRPHAAPPAAY  
 PHRNKVRAAEARTSAVRREYEAGDEASGRRGRGGRMKSVQNRADRGMWVRERRAVRGQENTVVVFAAA  
 ASPESDFLOIVLRDIAREMVPEYKSDLQPRQMDAKKPRKCDLTPFLVLKARKKQKFTSAKHLWVQRKGKG  
 VEDHADWQGIWVGVEGKSAEGQGLNLGTAFNRSPENKVSPOGPALNALEIDSSYLDPOALALPQGWQGLL  
 FPLLGTLELVGSKLSVQIQKPPSNIKNSRMTQVFHKNSTSVTSLPFVDTKGKKNVTSFPHISKVLLKSSLL  
 YQVSADGLDPEKAGERPGFSPKPVLERPRIVGKSTVAAEEENQAHNQMPASELKASEIPFHPSIKTQDPKA  
 EEKSPKKQKVTLTAAEALKLFKNQLSPYEQSEILGYAELWFLGLEAKKLDTAPEKFSKTSFDDHGFYLVK  
 LHDHIAIYRYEVLETIGKGSFGQVAKCLDHKNNELVALKIIRNKKRFHQQALMELKILEALRKDKDNTYNV  
 VHMKDDFFYFRNHFCITFELLGINLYELMKNNNFQGFSLIVRRFTLSVLKCLQMLSVEKIIHCDLKPENIV  
 LYQKGQASVKVIDFGSSCYEHQKGHLPSRLRSVTTTTSMILNKVLRSEPKLPHLQFEKVVSKY

>SGK295\_KIS\_ID#AA\_91

MAGSGCAWGAEPFLEAFGRWLQVQSRGSGSSASVYVRVCCGNPGSPPGALKQFLPPGTTGAAASAAEY  
 GFRKERAALQEQGHRNIVTLYGVETIHFSPNVPSRCLLLELLDVSVSELLLYSSHQGCMMWMIQHCARDV  
 LEALAFLLHHEGYVHADLKPRNILWSAENECFKLIDFGLSFKEGNQDVKYIQTGYRAPEALQNCQAQGL  
 QSDTECTSAVDLWSLGIILLEMFGMKLKHTVRSQEWKANSSAIIDHIFASKAVVNAIIPAYHLRDLIKSM  
 LHDDPSRRI PAEMALCSPFFSIPFAPHIEDVLMPTPVLRLNLVLDLDDYLENEEYEDVVEDVKECQKYG  
 PVVSLLVPGKNPGRGQVVFVEYANAGDSKAAQKLLTGRMFDGKFVVATFYPLSAYKRGYLYQTL

>SGK419\_ID#AA\_92

MKKFSRMPKSEGGSGGGAAGGGAGGAGAGCGSGGSSVGVVRFVAVGRHQVTLLEESLAEGGFSTVFLVRTH  
 GGIRCALKRMYVNNMPDLNVCKREITIMKELSGHKNIVGYLDCAVNSISDNVWEVLILMEYCRAGQVNVQM  
 NKKLQTGFTEPEVLQIFCDTCEAVARLHQCKTPIIHRDLKVENILLNDGGNYVLCDFGSATNKLNPQKDG  
 VNVVEEIKKYTTLSYRAPEMINLYGGKPITTKADIWALGCLLYKLCFFTLFPFGESQVAICDGNFTIPDNS  
 RYSRNIHCLIRFMLEPDPEHRPDIQVSYFAFKFAKKDCPVSNINKCCKQLLRHGALLTEILLFLQLFLNR  
 MTASEAAARKSQIKARITDTIGPTETSIAPRQRPKANSATTATPSVLTIQSSATPVKVLAPGEGFNHRPKG  
 ALRPGNGPEILLGQGPQQPPQOHRVLQQLQOQGDWRLQQLHLQHRHPHQOQOQOQOQHHLHLLQD  
 AYMQQYQHATQOQOQMLQOQFLMHSVYQPPQPSASQYPTMMPQYQQAFFQOQMLAQHQPSQQQASPEYLTSPQ  
 EFSPALVSYTSSLPAQVGTIMDSSYSANRSVADKEAIANFTNQKNISNPPDMSGWNPFGEENFSKLTEEEL  
 LDREFDLLRSSKGHLKAYFASQ

>SGK125\_MY03A\_ID#AA\_93

MFLIGKTIIFDNFPDPSDTWEITETIGKGTYGKVFVKNKNGQKAAVKILDPIHDIDEIEAEYNILKA  
 LSDHPNVVRFYGIYFKKDKVNGDKLWLVLELCSGGSVTDLVKGFLLKRGERMSEPLIAYILHEALMGLQHLH  
 NNKTIHRDVKGNILLTTEGGVKLVDFGVSAQLTSTRHRNRTSVGTFWMAPEVIACEQQLDTTYDARCDT

Figure 2G

WSLGITAIELGDGDPPLADLHPMRALFKIPRNPPLKRLQPELWSAEFNDFTSKCLTKDYEKRP<sup>T</sup>VSELLQH  
 KFITQIEGKDVMLQKQLTEFIGIHQCMGGTEKARRERIHTKKGNFNRPLISNLKDVDDLATLEILDENTVS  
 EQLEKCYSRDQIYVYVGDILIALNPFQSLGLYSTKHSKLYIGSKRTASPPHIFAMADLGYQSIITYNSDQC  
 IVISGESGAGKTENAHLLVQQLTVLGKANNRTLQEKILQVNNLVEAFGNACTIINDNSSRFGKYLEMKFTS  
 SGAVVGAQISEYLLKSRVHQHQAIGKNFHFYIYAGLAEKKKLAHYKLPENKPPRYLQNDHLRTVQDIM  
 NNSFYKSQYELIEQCFKVIGETMEQLGSIYSILAAILNVGNIEFSSVATEHQIDKSHISNHTALENCVSLC  
 CIRADELQEAALTSHCVVTRGETIIRPNTVEKATDVRDAMAKTLYGRLFSWIVNCINSLLKHDSSPSGNGDE  
 LSIGILOIFGFENFKKNSFEQLCINIANEQIYYYNQHVFAWEQNEYLNEVDVARVIEYEDNWPLDMFLQ  
 KPMGLLSLLDEESRFPKATDQTLVEKFEGNLKSQYFWRPKRMELSFGEIHHYAGKVLYNASGFLAKNKTLP  
 TDIVLLLRSSDNSVIRQLVNHPLTKTGNLPHSKLMLNYQMRTSEKLINLAKGDTGEATRHARETTNMKTQ  
 TVASYFRYSIMDLLSKMVVGQPHFVRCIKPNSEKQARKYDKEKVLLQLRYTGILETARIRRLGFSHRILFA  
 NFIKRYLLCYKSSEPRMSPDTCATILEKAGLDNVALGKTKVFLKYYHVEQLNLMRKEAIDKLILIQACV  
 RAFLCSRRYQKIQEKRESAIIQSAARGHLVGKQKEIVDMKNTAVTTIQTSDQEFDYKKNFENTRESFV  
 KKQAEANAIANERFISAPNNKGSVSVVKTSTFKPEEETTNAVESNNRVYQTPKKMNNVYEEVEKQEFYLVG  
 PEVSPKQKQFVKDLEENSNLRKVEKEEAMIQSYQRYTEERNCEESKAAYLERKAISERPSYPVWLAENET  
 SFKKTLEPTLSQRSIYQANANSMEKEKTSVVTQSAPICSQEEGRRLRHETVKERQVEPVTAQEEEDKAA  
 VFIQSKYRGYKRRQQLRKDKMSSFKHQKIVTTPTEVARNTNHLYSYPTKHEEINNIKKKDNKDSKATSERE  
 AGLAIFSKQISKLSEYFILQKKLNEMILSQQLKSLYLGVSHHKPINRRVSSQQLSGVCKGEEPILRP  
 PRRPRKPKTLNPNPEDSTYYYLLHKSQIEEKRRPRKDSQGLLDLEDFYKFEPLSRSGPKHSPSLRERRP  
 QQELQNCQIKANERCWAAESPEKEEREPAANPYDFRRLLRKTSQRRRLVQQS

>SGK445\_ID#AA\_94

FKVGNLPKGSFAGVCKVESIHTGLEISVQMIVKKAVRIQRVHNEMKMHQCQLKHPPTLELYKYFKDKD

>SGK127\_ID#AA\_95

MTKSEEQQLSLQKALQQCELVQNMIDLSISNLEGLRTKCATSNDLTQKEIRTLESKLVKYFSRQLSCKKK  
 VALQERNAELDGFQPLRHWFRIVDVRKEVLEEISPGQLSLEDLLEMTDEQVCETVEKYGANREECARLNAS  
 LSCLRNVMHSGGNLSKQDWTIQWPTTETGKKNPVCPEPTPWIRTHLSQSPRVPSKCVQHYCHTSPTPGA  
 PVYTHVDRLTVDAYPGLCPPLPESGHRSLPPSPRQRHAVRTPPRTPNIVTVPPTPPMRKKNKLPFG  
 TPPPSSRKLHLIPGFTALHRSKSHFQLGHRVDEAHTPKAKKKSKPLNLKIHSSVGCENIPSQQRSPLL  
 SERSLRSFFVGHAFLPSTPPVHTEANFSANTLSVPRWSPQIPRRDLGNSIKHRFSTKYWMSQTCTVCCKG  
 MLFGLKCKNCKLKCHNKCTKEAPPCHLLIHRGDPARLVRTESVPCDINNPLRKPFRYSDLHISQTLPKTN  
 KINKDHIIPVYQPDSSSNPSSTTSSTPSSPAPLPSPSATPPSPPLHPSPOCTROQKNFNLPAHYKYKQOF  
 IFPDVVPVPEPTPAPQVILHPVTSNPILGNPLQIEVEPTSENEEVHDEAESEDDFEEMNLSLLSARS  
 FPRKASQTSIFLQEWDPFEQLEIGELIGKRGFQVYHGRWHGEVAIRLIDIERDNEDQLKAFKREVMAYR  
 QTRHENVVLFMGACMSPPHLAIITSLCKGRTLYSVVRDAKIVLDVNKTRQIAQEIIVKGMGYLHAKGILHKD  
 LKSKNVFYDNGKVVITDFGLFSISGVLAGRRREDKLRIQNGWLCHLAPEIIRQLSPDTEEDKLPFSKHSV  
 FALGTIWEYELHAREWPFKTQPAEAIWQMGTMKPNLSQIGMGKEISDILFCWAFEQEERPTFTKLMML  
 EKLPRNRRLSHPGHFWKSAL

>SGK009\_ANRD3\_ID#AA\_96

MEGDGGTPWALALLRTFDAGEFTGWEKVGSGGFGQVYKVRHVHWKTWLAIKCSPSLHVDDRERMELLEAK  
 KMEMAKFRYILPVYGICREPVLVMEYMETGSLEKLLASEPLPWDLRFRIIHETAVGMNLFHCMAPLLHL  
 DLKPANILLDAHVKISDFGLAKCNGLSHSHDLSMDGLFGTIAYLPPERIREKSRLFDTKHDVYSFAIVI  
 WGLTQKKPFADEKNILHIMVKVVGHRPELPPVCRRARPRACSHLIRLMQRCWQGDPRVRPTFQGNGLNGE  
 LIRQVLAALLPVTGRWRSPGEGFRLESEVIIIRVTCPLSSPQIEITSETEDLCEKPDDEVKETAHDLVDKSP  
 EPRSEVVPARLKRASAPTFDNDYSLSELLSQLDGVSQAVEGPEELSRSSSESKLPSGSGKRLSGVSSVD  
 SAFSSRGSLSLSFEREPSTSDLGTTDVQKKLVDAIVSGDTSKLMKILQPQDVLALDSGASLLHLAVEAG  
 QEECAKWLNNANPNLSNRRGSTPLHMAVERRVRGVVELLLARKISVNAKDEDQWTALHFAAQNGDESST  
 RLLLEKNASVNEVDGRTPMHVACQHGQENIVRIILLRGVDVSLQGKDAWLPLHYAAWQGHLPVIVKLLAK  
 QPGVSVNAQTLDGRTPLHLAAQRGHRVARILIDLCSNVNCSLLAQTPHVAETGHTSTARLLLHRGAG  
 KEAVTSDGYTALHLAARNGLHATVKLLVEEKADVLARGPLNQALHLAAAHGHSEVVEELVSADVIDLDFDE  
 QGLSALHLAAQRHAQTVETLLRHGAHINLQSLKFQGGHGAATLLRRSKT

>SGK421\_STK22A\_TSK1\_ID#AA\_97

MDDAAVLKRRGYLLGINLGEYSYAKVKSAYSERLKFNVAIKIIDRKKAPADFLEKFLPREIEILAMLNHCS  
 IIKTYEIFETSHGKVYIVMELAVQGDLELIKTRGALHEDEARKKFHQLSLAIKYCHDLVDVHRDLKCDNL

Figure 2H

LLDKDFNIKLSDFSFSKRCLRDDSGRMALSKTFCGSPAYAAPEVLQGIPIYQPKVYDIWSLGVILYIMVCGS  
MPYDDSNIKKMLRIQKEHRVNFPRSKHLTGECCKDLIYHMLQPDVNRRLHIDEILSHCWMQPKARGSPSVAI  
NKEGESSRGTEPLWTPEPGSDKKSATKLEPEGEAQQAQPETKPEGTAMQMSRQSEILGFPSKPSMETEE  
GPPQQPPETRAQ

>SGK047\_ID#AA\_98  
PREAVSIVSSCWVRPPATRLRALKILTHPAF

>SGK196\_ID#AA\_99  
MEKQPQNSRRGLAPREVPPAVGLLLIMALMNTLLYLCLDHFFIAPRQSTVDPTHCPYGHFRIGQMKNCS  
LSCEELRTEVRQLKRVGEGAVKRVFLSEWKEHKVALSQTLSLEMKDDFLHGLQMLKSQGTHTVTLG  
DDNTMLTEYHPLGSLSNLEETLNLKYQNVNTWQHRLELAMDYVSIINYLHHSVPVGTVMCDSDNLPK  
QYLLTSNFSILANDLDALPLVNHSSGMLVKCGHRELHGDFVAPEQLWPYGEDVPFHDDLMP  
IPDISSFLLGHIEGSDMVRFLFDIHKACKSQTPSERPTAQDVLETYQKVLDTLRDAMMSQAREML

>SGK396\_ID#AA\_100  
MDNIDEILEKTESSVCKELEIALVDQGDADKEIISNTYSQVLQKIHSEERLIATVQAKYKDSIEFKKQ  
YLKKIPSVDHLLSIKKTLSLKLALRWKLVEKSNLEESDDPDGSQIEKIKEEITQLRNNVFQEIYH  
EMLTSLAQKWPELPLHPEIGLLKYMNSGGLLTMSLERDLDLDAEPMKELSSKRPLVRSEVNGQI  
SVDVDTEAKVIERAATYHRAWREAEGDGLLPLIFLFLCKSDPMAYLMVPPYPRANLNAVQANMPL  
LKVMKGVAQGLHTLHKADIHGSLSHQNNVFALNREQGIQVDFDFTKSVSQRASVNMVGDLSLMS  
KPASPGSDLYAYGCLLLWLSVQNOFEINKDGIKVDQFHLDDKVKSLLCSLICYSMTAEQVLNAE  
MPKEQSVNPEKDTETLYKKEEIKTENLDKCKMEKTRNGEAFDC

>SGK279\_PKN\_ID#AA\_101  
MSVGCPEPEPRSLTCCGPGTAPGPGAGVPLLTEDMQALTIRTLAASDVTKHYELVRELKGT  
YKGTGTMALKFVNKSKTKLKNFLREVSITNSLSSSPFIKVFDDVFETEDCYVFAQYAPAGDL  
QVGLPEDTVKRCVQQLGLALDFMHGRQLVHRDIKPNVLLFDRECRVKLADFGMTRRVGCRV  
YTAPEVCQAGRADGLAVDTGVDVWAFGVLIFCVLTGNFPWEAASGADAFEEFVRWQGRGL  
TEPALRMFORLLALEPERRGPAKEVFRFLKHELTSELRRRPSHRARKPPGDRPPAAGPLRLE  
LTESGGGSRPAPPVGSVPLVPVPVPVPVPVPVPEPGLAPQGPPGRDGRADKSKQVVLATAIEICV

>SGK037\_ID#AA\_102  
MDKYDVIKAIGQAFGKAYLAKGKSDSKHCVIKEINFEKMPIQEKEASKKEVILLEKMKHPNIVAF  
ENGRFLFIVMEYCDGGLMKRINRQGVLFSEDQILGWVQISLGLKHIHDRKILHRDIKAQNI  
AKLGDFGIARVLNNSMELARTCIGTPYYLSPEICQNKPYNNKTDIWSLGCVLIELCTLKHPFEG  
LKICQAHFAPISPGFSRELHSLISQLFQVSPDRPSINSILKRPFLNLIPKYLTPEVSFEVTVW  
GCRSLTRVFGFRCEPLHPACIMVFKNIPSCFLQIKMIERPPIAAVCCHYDYAQLDMLRRRAH  
PQENTGVEDYGQETRHHGSPSPQWPAEYLQRKFEAQYKLVKQGLRPSSAEPNYNQRELRSNGE  
QELPFRKNEMKEQEYWKQLEEIRQQYHNDMKEIRKKMGREPEDIKDLQMRLONTKESKNPEQY  
VKFEINLDKCISDENILQEEEAAMDIPNETLTFEDGMKFKEYECVKEHGDYTDKAFEKLHCPE  
AVGNRRQWDGGAPQTLQMMAVADITSTCPTGPDNGQVIVIEGIPGNRKQWRHEAPGTLMSVLA  
SFSADDEEFGM

>SGK060\_ID#AA\_103  
MLKFQEAACVSGSTAISTYPKTLIARRYVLQQLGSGSFGTVYLVSDKKAKRGEELKVLKEISV  
ETVQANLEAQLSKLDHPAIVKFHASFVEQDNFCIITEYCEGRDLDDKIQEYKQAGKIFPENQI  
LLGVDMHERRILHRDLKSKNVFLKNNLLKIGDFGVSRLLMGSCDLATTLTGTPHYMSPEALKH  
DIWSLACILYEMCCMNAFAGSNFLSIVLKIVEGDTPSLPERYPKELNAIMESMLNKNPSLRPSA  
PYLDEQLQNLNCRYSEMTLEDKNLDCQKEAAHIINAMQKRIHLQTLRALSEVQKMTPRERMRLR  
KARKLKKIVEEYKSKRMQELRSRNFQQLSVDVLHEKTHLKGMEKEEQPEGRSCLSPQDEDEER  
EESDEPTLENLPESQPIPSMDLHELESIVEDATSDLYHEIPEDPLVAEEYADAFDSYCVESDE  
ALERPEKEIRNEGSQPAYRTNQDSDIEALARCLENVLGTSLDTKITTTMAEDMSPGPPIFN  
KRMRESAMQKLGEVFEVYNYLKRARHQNAEAEIRECLEKVVPQASDCFEVDQLLYFEEQLITMKEP  
TLQNLH

>SGK080\_ID#AA\_104

Figure 21

MPSRAENYEVLYTIGTSGRCQKIQRKSDGKILVWVWELHYGSMTEAEKQMLVSEVNLCELKNPNIVHYH  
DRIIDRTNTTLYIVMEYCEEGLASVITKGTKERQYLDEEFVLRVTQTALALXCHRRSDGDHTVVRDL  
KPASVFLDGKQNVKLGDLGLARILNHDTSFAKTFVGTPTYMSPEQTNHMSYNEKPDIVSLGCLPYESRALM  
PPFTAFSQKELAGKIREGKFRILYRYSDELNEIMRMLKDYHRPSVEEILENPLIADLVAEEQRRNLERR  
GRQLGEPEKLPDSSPVLSELKLKEIQLEEQERALKAGEERLEQKEQELCVCERLAEDRLAI PENLLKNYSL  
LKEQKFLSLASSPELLNLPSSVIKKKVHFSGESKENVMRSENPEQLTSKSKCKDLKXXCFMLHAAQLRAQ  
ALSDIEKNYQLKSRQILGMR

>SGK002\_ID#AA\_105

MLARRKPMPLALTINPTIAEGPSPTSEGASEANLVDLQKKLEELDEQQKRLAEFLTQKAKVGELKDDDF  
ERTSELDAGNGGVVTKVQHRPSGLIMARKLIHLEIKPAIRNQI IREHQVLHECNSPYIVGFYGFYCDREI  
SICMEHMDGGSLDQGLKEAKRIPEDILGKVSIAVLRGLAYLREKHQIMHRNVKPSNIVNSRGEIKLCDFG  
VSGQLIDSMANSFVGTSTRYMAPERLQGTHTSVQSVIWSMDLSLVELAIERYPIPPPDAAKELEAIFGQPVVD  
REEGEPHSISSWPGSPGRPNSSGYGMDSLPAMAI FELLDYIVKEPPPKLPNGVFTPDFQEFVNKCLIKNPTÉ  
RADLKMILTINHAFIKRSEVKEADFACXLCCTLGLNQPGTPTRTAV

>SGK058\_ID#AA\_106

NSLKSEEPILWTKGEILGKAYGTVYCGLTSQGQLIAVKQVALDTSNKLAAEKEYRKLQEEVDLLKALKHV  
NIVAYLGTCLQENTVSIFMEFVPGGSISSIINRFGLPPEMVFCYTKQILQGVAYLHENCVVHRDIKGNV  
MLMPTGIIKILIDFGCARRLAWAGLNGTHSDMLKSMHGTPYWMapeVINESGYGRKSDIWSIGCTVFEMATG  
KPPLASMDRMAAMFYIGAHRGLMPPLPDHFSENAADFVRMCLTRDQHERPSALQLLKHSFLERSH

>SGK103\_ID#AA\_107

DLTPQNILLDKSVFVYLGVLAREFML

>SGK035\_ID#AA\_108

RDCPPVIAPPDHMKSIYTRAVIDPVLAPVGDSNVDGGAKSLDKQKKTKMTDEEIMEKLRTIVSIGPVRK  
KYTRYEKIGQASGTVFTATDVALGQKVAIKQINLQKQPKKELI INEILVMKELKNPNIVNFLDSYLVGDE  
LFVVVEYLARGSLTDVVTETCMDEAQIAAVCRESLQALEFLHANQVIHROIKSDSVLLGMEGSVKLTDFGF  
CAQITPEQSKRSTVVRTPHWMAPEVVTRKAYGPKVSVWSLGIMATEMVEGEPPYLNENPLRALCLTIATNGI  
PELQNPETLSPIFRDFLNRCLTDVEKRGSAKELLQHLFLKLAKLLSSLTPLIMAAKEAMKSNR

>SGK075\_ID#AA\_109

MMGLGESLPDPTDWEIIETIGKTYGKVYKVTNKRDSGLAAVKILDPVSDMDEEIEAEYNIQFLPNHPN  
VVKFYGMFYKADHCVGGQLWLVLLELCNGGSVTELKGLLRCCQRLDEAMISYILYGALLGLQHLHNNRIH  
RDVKGNNILLTTEGGVKLVDFGVSAQLTSTRLRNRTSVGTFFWMAPEVIACEQQYDSSYDARCDVWSLGIT  
AIELGDGDPPLFDMHPVKTLFKIPRNPPTLLHPEKWCEEFNFHIFISQCLIKDFERRRPSVTHLLDHPFIKGV  
HGKVLFLQKQLAKVLQDQKHQNPVAKTRHERMHTRRPYHVEDAEKYCLEDDLNVNLEVLDEDTIIHQLOKRY  
ADLLIYTYVGDIILALNPFQNLISIYSPQFSRLYHGKVRASNPPIHIFASADAAYQCMVTLTKDQCIVISGES  
GSGKTESAHLIVQHLTFLGKANNQTLREKILQVNSLVEAFGNSCTAINDNSSRFKYLEMMFTPTGVVMGA  
RISEYLLKSRVIKQAAREKNFHI FYIYIYAGLHHQKLSDFRLPEEKPPRYIADETGRVMHDITSKESYRR  
QFEAIQHCFRIIGFTDKEVHSVYRILAGILNIGNIEFAAISSQHQTDKSEVPNAEALQNAASVLCISPEEL  
QEALTSHCVVTRGETIIRANTVDRAADVDRAMSKALYGRFVSWIVNRINTLLQPDENIWQVPRRAEGGGMN  
VGILDIFGFENFQNSFEQLCINIANEQIYFYFNQHVFALEQMEYQNEGIDAI PVEYEDNRPLDMFLQKP  
LGLLALLDEESRFQATDQTLVDKFEDNLCKYFWRPKGVLCFGIHYAGKVLYDASGVLEKNRDTLPAD  
VVVVLRTSENKLLQQLFSIPLTKGTWEPSPDPAKAFARPGNLAQTRARITVASSSLPPHFSAGKAKVDT  
LEVIRHPEETTMMKRQTVASYFRYSLMDLLSKMVGQPHFVRCIKPNDREALQFSRERVLAQLRSTGILE  
TVSIRRQGYSHRILFEEFVKRYYYLAFTAHQTPLASKESCAVAILEKSRLDHWVLGKTKVFLKYYHVEQLNL  
LLREVIGRVVVLQAYTKGWL GARRYKKVREKREKGAIAIQS

>SGK188\_EPHA9\_ID#AA\_110

METCAGPHPLRLFLCRMQLCLALLGPWRPGTAEVILLDSKASQAEWGWTALPSNGWEEISGVDEHDRPI  
RTYQVCNVLEPNQDNWLQGTGWSRGRGQRI FVELQFTLRDCSSIPGAAGTCKETFNVYYLETEADLGRGRP  
RLGGSRPRIKIDTIAADESFTQGDGGERKMKLNTTEVREIGPLSRRGFHLAFQDVGACVALSVRVYKQCRA  
TVRGLATFPATAESAFASTLVEVAGTCVAHSEGEPSPPRMHCGADGEWLVPVGRSCSAGLFQERGDICEA  
CPPGFYKVSPPRRVCSPCPEHSRALENASTFCVCQDSYARSPTDPPSASCTRGPPSAPRDLQYSLSRSPLV  
LRLRLWLPADSGGRSDVTYSLLCLRCGREGPAGACEPCGPRVAFPRQAGLRERAATLLHLRPGARYTVRV

Figure 2J

AVLNGVSGPAAALVPVGAVSINPGTVGVPVAGVIRDRVEPQSVSLSWREPAPAGAPGANDTEYEIRYYEK  
VQSEQTYSMVKTGAPTVTVTNLKPATRYVFQIRAAASPGPSWEAQSFNPSIEVQTLGEAASGRDQSPAIVV  
TVVTISALLVLGSVMSVLAIWRRPCSYGKGGGDAHDEEELYFHFVKVTRRTFLDPQSCGDLQAVHLFAKE  
LDAKSVTLERSLGGGRFGELCCGCLQLPGRQELLVAVHMLRDSASDSQRLGFLAEALTGQFDHSHIVRLE  
GVVTRGSTLMIVTEYMSHGALDGLRRHEGQLVAGQLMGLLPGLASAMKYLSEMGYVHRGLAARHVLVSSD  
LVCKISGFGGRPRDRSEAVYTTMSGRSPALWAAPETLQFGHFSSASDVWSFGIIMWEVMAFGERPYWDMMSG  
QDVIKAVEDGFRLLPPRNCNPNLLHRLMLDCWQKDPGERPRFSQIHSILSKMVQDPEPPKCALTTCPRPPTP  
LADRAFSTFSPFGSVGAWLEALDLCRYKDSFAAAGYGSLEAVAEMTAQRDLVSLGISLAEHREALLSGISA  
LQARVLQLQGQGVQV

>SGK040\_ID#AA\_111

MQPLDFSSGSDPNISLSEKIRDQLVVGQILPDCYVELEKIIILSERKNVPIEFVVIDRKRLLQLVRENQLQ  
LDENELPHAVHFLNESGVLLHFQDPALQLSDLYFVEPKWLCKIMAQDVSSIFGLYIRDILTVEGCPKHP  
KGIISRRDVEKFLSKKRKFKNYMSQYFKLLEKFQIALPIGEEYLLVPSSLSDRPVIELPHCENSEIIIR  
LYEMPYFPMGFWSRLINRLLLEISPYMLSGRGICLLGQVVDHIDSLMEWFPGLLEIDICGEGETLLKKWAL  
YSFNDGEEHQKILLDDLKKAEEGDLLVNPDPRLTIPISQIAPDLILADPPRNIMLNDELEFEQAPEFL  
LDCFVCIHLYPSSDYISRHYMRTINIVQTGFAKRWRTVHGADHGDGSGFSVYRAAYEGEEVAVKIFNKH  
TSLRLLRQELVVLCHLHHPISLISLLAAGIRPRMLVMEASKGSLDRLLQODKASLRTLQHRIALHVDGL  
RYLHSAMIIYRDLKPHNVLLFTLYPNAIIAKIADYGIAQYCCRMGIKTSEGTGFRAPAEVARGNVIYNQQ  
ADVVSFGLLLYDILTGGRIVEGLKFPNEFDELEIQKLPDPVKEYGCAPWPMVEKLIKQCLKENPQERPT  
SAQVFDILNSAELVCLTRRILLPKNVIVECMVATHHNSRNASIWLGCCHTDGQSLFOLDNTEGYTSEEVA  
DSRILCLALVHLPVEKESWIVSGTQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVG  
TADGKLAI FEDKTVKLGAAPLKILNIGNVSTPLMCLSESTNTERNVMWGGCGTKIFSFNDFTIQKLE  
TRTSQLFSYAAFSDSNIITVVVD TALYIAKQNSPVVEVWDKKEKLCGLIDCVHFLR

>SGK390\_ID#AA\_112

MAGAGQHHPGAAGGAAAGAGAAVTSAAASAGPGEDESSEAEQEGPQKLIRKVSTSGQIRTKTSIKEGQ  
LLKQTSFQRWKKRYFKLRGRTLYYAKDSKSLIFDEVDLSDASVAEASTKNANNSFTIITPFRRMLCAEN  
RKEMEDWISSLSKVQTPREPYEAQFNVEHFSGMHNWYACSHARPTFCNVCRESLSGVTSHGLSCEVCKFKA  
HKRCAVRATNNCKWTTLASIGKDIIEDEGDVAMPHQWLEGNLPVSAKCAVCDKTCGSVLRQLQDWKCLWCKT  
MVHTACKDLYHPICPLGQCKVSIIPPIALNSTDSGFCRATFSFCVSPLLVFVNSKSGDNQGVKFLRRFKQ  
LLNPAQVFDLMNGGPHLGLRLFQKFDNFRILVCGGDGSGVGVWLSEIDKLNLNKQCQLGVLPGLGTGNDLARV  
LGWGGSYDDDTQLPQILEKLERASTKMLDRWSIMTYELKLPPKASLLPGPPEASEEFYMTIYEDSVATHLT  
KILNSDEHAVISSAKTLCETVKDFVAKVEKTYDKTLENNAVADAVASKCSVLNEKLEQLLQALHTDSQAA  
PVLPLSPLIVEEDAVESSESSESLGESKEQLGDDVTKPSSQKAVKPREIMLRANSLKKAQRVIEEAGKVM  
DDPTVHPCEPANQSSDYDSTETDESKEEAKODGAKESITVKTA PRSPDARASYGHSQTDVSPGPAVAASKE  
NLPVLNTRIICPGLRAGLAASIAGSSIINKMLLANIDPFGATPFIDPDLSDVDGYSEKVMNNYFGIGLDA  
KISLEFNNKREEHPEKCRSRTKNLMWYGVLTRELLQRSYKNLEQRVQLECDGQYIPLPSLQGI AVLNIPS  
YAGGTNFWGGTKEDDIFAAPSFDDKILEVVAIFDSMQMAVSRVIKQHHRIAQCRTVKITIFGDEGVVPVQV  
DGEAWVQPPGIIKIVHKNRAQMLTRDRAFESTLKSWEKQKCDSGKPVLRTHLYIHHAI DLATEEVSQMQL  
CSQAAELITRICDAATIHCLLEQELAHAVNACSHALNKANPRCPESLTRDTATEIAINVKALYNETESLL  
VGRVPLQLES PHEERVSNALHSVEVELQKLT EIPWLYIILHPNEDEEPPMDCTKRNNRSTVFRIVPKFKE  
KVQKQKTSSQPGSGDTESGSCEANSPGN

>SGK007\_ID#AA\_113

MALKPCLEAPIESGLCSGIECHAEHPTLVLM LFASVLVTCLEAAKLTVGFTQTPWNISHPFMSQRLGAGLOI  
AMDKVNSELVDLGNFTTYTNSACSTKESLAI FNNQVQNEQISALFGPACPEAAEDIGDVLQESLQYLGWKH  
IGMFGGYSGASSWDGVDELWRVVENELKSHFIIITASMRYTNNSLVLLQEH LWRISSIARV IITCSSEDAK  
IILLAAANLGLNTGEFVFIILQQLLEVGTVDHSMALVVRGPFLEGSTDKSENDAPPKGVVWTASHCPQLLQK  
RPWRRRLLETSLPNTTEVALPQHLLGAGGCQDLDPGWSGQRCGARTGQGGQVSPYSAYLHDAVLLYAET  
VKQVVKAGGDFQDGWQLVSALKGSSQTTVQGITGPVFVDAQGERHMDYSVYALQKSENGPLLSFLHYDSY  
QSVTAVLLTLMILIPVLGAAIIGLILRMQRQNKDIWWQINFDDITILPQNKPSQRATPVSKGINSNSSVM  
ISVDLSSFVKSQWEELFYAAVGLYQGNHVAIRYVGDAEAWVRKPIVLQEIQLLAAYTFTTRSDKASDLG  
VSPGMLFLHRSPLGSHSNLKPSNCLMDGRLQYPHSMRTSHSFAELYWTAPELLQFPMPWSGTPQGDVYSF  
AILMRELIYHWDHGPFDLHEAPDEIINQIKDPAAVPLQPSLPEEKGNEKIVAMVRVCWDESLEKRPFSF  
SIKKTLEASPRGHVSILASVMSKLEVYANYLEEUVQERTSQLTAEKRKVEKLLSTKVPSTGEQQLAGRS

Figure 2K

VEPEHFESVTIFLSDIVGFTKLCSLSSPLQVVKLLNDVYSLFDHIITTYDVYKGKGEQTTFWLKDEGFTL  
PLPNLLRKKPKSQRYCELISLQGDRSSYLTAS

>SGK050\_ID#AA\_114

RELDHPNCRFTGGCIALPDVVIVMEYCPKGSMLMDVLLNDNISFNWGF

Figure 2L

**BEST AVAILABLE COPY**



(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
31 May 2001 (31.05.2001)

PCT

(10) International Publication Number  
WO 01/38503 A3

(51) International Patent Classification<sup>7</sup>: C07K 14/47.  
C12N 9/12, A61K 38/45, C07K 16/40

(71) Applicant (for all designated States except US): SUGEN, INC. [US/US]: 230 East Grand Avenue, South San Francisco, CA 94080 (US).

(21) International Application Number: PCT/US00/32085

(22) International Filing Date:  
22 November 2000 (22.11.2000)

(72) Inventors; and  
(75) Inventors/Applicants (for US only): PLOWMAN, Gregory, D. [US/US]: 4 Honeysuckle Lane, San Carlos, CA 94070 (US). WHYTE, David [US/US]: 2623 Barclay Way, Belmont, CA 94002 (US). MANNING, Gerard [US/US]: Sugen, Inc., 230 East Grand Avenue, South San Francisco, CA 94080 (US). SUDARSANAM, Sucha [US/US]: Sugen, Inc., 230 East Grand Avenue, South San Francisco, CA 94080 (US). MARTINEZ, Ricardo [US/US]: 984 Cartier Lane, Foster City, CA 94404 (US). FLANAGAN, Peter [US/US]: 192 Liberty Street, San Francisco, CA 94110 (US). CLARY, Douglas [US/US]: Sugen, Inc., 230 East Grand Avenue, South San Francisco, CA 94080 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/167,482 24 November 1999 (24.11.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/167,482 (CON)  
Filed on 24 November 1999 (24.11.1999)

[Continued on next page]

(54) Title: NOVEL HUMAN PROTEIN KINASES AND PROTEIN KINASE-LIKE ENZYMES

>SGR15\* CK15 IC BNA 3  
ATGTTGAAGTTCAAAATATGGAGCGCGGAATCCTTGGATGCTGGTGTGACCAACCATTCGCCAGCCGGGC  
CTCCAGGCTGAACTCTCTCTCCAGGGGAAACCCCTTTATGACTCAACAGCAGATGTCTCTCTTTCCG  
GAGAAAGGATATTAGATGCTCTCTTTGTTCTCTTTGAAGAATGCACTGAGCTGTCTGATGAAGATTAG  
CACGTGAGGAACTTTCTCCGGAAGTGTATCCGACACCATAGCTGAGTTACAGGAGCTCCAGGCTTCGGC  
AAAGGACTTCGAAGTCAGAAATCTTGTAGGTTGTGGTCACTTTGCTGAAGTGCAGGTGGTAAGAGAGAAAG  
CAACCGGGACATCTATGCTATGAAAGTGAAGAAGAAGGCTTTATTTGGCCAGGAGCAGGTTTCATT  
TTTGAAGAGAGCGGACATATATCTCGAAGCACAAGCCGCTGGATCCGCCAATTACAGTATGCTCTTCA  
GGACAAAATTCACCTTATCTGGTCAATGGAATATCAGCTGGAGGGGACTTGTCTGCTCACTTTTGAATAGAT  
ATGAGGACCACTAGATGAAACCTGATACAGTTTACCTAGCTGAGCTGATTTTGGCTGTTCACAGCGTT  
CATCTGATGGGAACGTCGATCGAGACATCAAGCTTGAGAACATTTCTGTTGACCGCAGAGGACACATCAA  
CTGCTGGATTTTGGATTCGCGCGGAAAAATGAATTCAAACAAGTGGTGAATGCCAAATCCCGATTGGGA  
CCCCGATTACATGCTCTCTGAAGTGTGACTGTGATGAACGGGGATGAAJAGGCACCTACGGCTGGAC  
TTGACTGGTGGTCACTGGCTGATTCGCTATGAGATGATTTATGGAGATCCCTCTCCAGAGGGGAAC  
CTCTCCAGCAACCTTCAATAACATTATGAATTTCCAGCGGTTTGTGAATTTCCAGATGACCCCAAAGTGA  
GCATCACTTTCTTATGATGATTCAAAGCTTTGTTGGCGGCGAGAGAGAGACTGAAGTTTGAAGGCTCT  
TGCTGCTCTCTCTCTCTAAATTTGACTGGAAACATTCGTAACCTCTCTCCCTCCCTCTGCTTCCCA  
CTCTCAAGTTCAGCATGACACCTCCAAATTTGATGAACAGAGAGAAATTCGTTGGGTTTCACTCTCTCT  
GCCAGCTGAGCCCTCAGGCTTCTCGGGTGAAGAACTCCGCTTTGCGGTTTTCGTACAGCAAGGCACTG  
GGGATCTTTCTAGATCTGAGTCTGTGTGTGCGGCTTGGACTCCCTGCGCAAGACTAGCTCCATGGAAAA  
GAACTTTCTATCAAAAGCAAGAGCTACAGAGCTCTCAGGACAAAGTGTCAAGATGGAGCAGGAATGA  
CCCGGTTACATCGAGAGGTGTGAGAGGTGGAGGCTGTCTAGTCAGAGGAGGTGGAGCTGAAGGCTCT  
GAGACTCAGAGATCTCTCTGGAGCAGGACCTTCTACCTACATCAGAAATCAGTACGTTTAAAGCGAAG  
TTTGGAGCAAGTACGAGTGGAGGTGTCAGGAGGATGACAAAGCACTGACGCTTCTCCATGATATCAGAG  
AGCAGAGCCGCAAGCTCCAAGAAATCAAGAGCAGGAGTACAGGCTCAAGTGGAGAAATGAGGTTGATG  
ATGAATCAGTGGAGAGGATCTGTCTCTCAGCAAGAGCGGAGTGTCTTCAAGATCTGAGCTGAGAGA  
GTCTCGGCTGTCTCTCAAGAAATCAAGCGGAAGCGACAGAAATGTCAGCAATAAATTTTGAAGGCTAAGG  
ATCAAGCGGAAGCTGAAGTGGAGAAATTCGGAACCTGGAGAGATCAATGCTGAGCAGCAGCTCAAAAT  
CAGGAGCTCAAGAGAACTGGAGAAAGCTGTAAAAGCCAGCAGGAGGCGCCAGGCTGCTCCAGAAAT  
CCGCGAGGCAAGAGCGAGCGAGGAGGCTGGAGAGCTGCGAGAACCCAGAGGATTTCTCTGAAGGCA  
TCAGAAAGAGCTGCTGGAGCTGAGGAACCGCCCAITCTCTGAGAAACAGGTAAGAGACTAGAGACC  
ATGAGGCTGAGAAACAGACTGAAGGATGACATCCAGCAAAATCCCAAGATCCAGCAGATGGCTGA  
TAAATTTCTGAGCTCGAAGAGAAATCATCGGAGGCGCAAGTCTCAGGCCAGCAGCTAGAGGTGCACTGA  
AACAGAAAGAGCACTATCAGAAAGATTAAGCTTTGGACAAATCAGATAAAGAAAGACCTGGCTGAC  
AAGGAGACACTGGAACATGATCCAGAGACAGGAGGAGGCGCCATGAGAGAGGCAAAATTTCTCAGCGA  
ACAGAGGCGATGATCAATGCTATGAGATTCAGATCAGATCCCTGGAAACAGGAGATTGGAAGCTGTG  
AAGCAATCAATTTGAGCAAAATAGCAGCTTTTATCCCAAGAGCAATGAAGGCGCAAGAGAGATGATT  
TCTCACTCAGGCAACAGAAATTTTACCTGGAGACACAGGCTGGAGGTTGGAGGCGCAAGCAACCGAAAT  
GGAGGAGCAGCTGGAGAGATCAGCTCAGCAAGAGCAAGTGAAGAAATCGGCTGCTGGAACTGGAGACAA  
GATTTCGGGAGGTCAGTCTAGAGCAGAGGAGGAGCAAGCTGGAGCTCAAGGCGCAGCTCAGAGGCTACAG  
CTCTCCCTCAGGAGCGGAGTACAGTGTGACAGGCTGCGAGGCTGCGAGGCGGCGCTGGAGAGCCAGCT  
TCGCGAGGCGAAGCAGAGCTGGAAGAGACACAGCAGAGCTGAAGAGGAGATCAAGGCACTCAGGCACT  
ATGAGATGAAATCCAGCGCAAAATTTGATGCTCTTGTGAACAGCTGTACTGTATCAAGACCTGGAGGAG  
CAGCTAAACAGCTGACCGGAGCAACGCTGAATCAACAAACAAACTTCTACTTGTCCAAACAACTCGA  
TGAGGCTCTGGCGCAACAGCAGAGATTGTCAACTCGGAAGTGAAGTGAACCTCTCCGCGGGAGATCA  
CGGAACCTGAGATGAGCTTACAGCGCAAGCAAGCAAGTGGAGGCTTGAAGACCACTGTGCACTCACTG  
GAGGACAGGCTATGATTTGGAGGCGCTTAAAGCAAGAGCTGTAGAAJAGAGCGGAGTGGAGGAGGCTG  
GAGGAGGCTCTGGTGTGAAGAAATCCAGTTTGAAGTGTGCGGCTTGAAGCTGCTGAGAGCTGAGAGATG  
CCGAGAAACAGAGAGGCGGAGAGCGGATCAGCGGATCAGCGGCTGCGGAGGCTGCGGAGGCTGGAGG  
AAGGAGCAACAGGCTGAGATTCTGCTCTGCGAGGCTCTCAAGAGCAGAAAGCTGAAGGCGGAGGCT  
CTCTGAGAGCTCAATGACCTGGAGAGAGATGCTATGCTTGAATGAATGCCCGAGCTTACAGCAGA  
AGCTGAGGCTGAACGAGGCTCAAAACAGAGGCTTCTGGAAGAGCAAGCAAAATTAACAGCAGAGATGGA  
CTGCAAGAAATCACTTTTCTGCTGACTCAAGGAGCTCAAGAGCTCTAGATGGGCTGATCTACTGAA

(57) Abstract: The present invention relates to kinase polypeptides, nucleotide sequences encoding the kinase polypeptides, as well as various products and methods useful for the diagnosis and treatment of various kinase-related diseases and conditions. Through the use of a bioinformatics strategy, mammalian members of the PTK's and STK's have been identified and their protein structure predicted.

WO 01/38503 A3



(74) **Agent:** BURROUS, Beth, A.; Foley & Lardner, 3000 K. Street, NW, Suite 500, Washington, DC 20007 (US).

patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

**Published:**

— with international search report

(88) **Date of publication of the international search report:**  
31 January 2002

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/32085

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 C07K14/47 C12N9/12 A61K38/45 C07K16/40

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

BIOSIS, EMBL, EPO-Internal

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DI CUNTO FERDINANDO ET AL: "Citron Rho-interacting kinase, a novel tissue-specific Ser/Thr kinase encompassing the Rho-Rac-binding protein citron." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 273, no. 45, 6 November 1998 (1998-11-06), pages 29706-29711, XP002170360 ISSN: 0021-9258 the whole document --- -/--	1-28

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

19 September 2001

Date of mailing of the international search report

22.10.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Schwachtgen, J-L

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/32085

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WEISS ELLEN R ET AL: "The cloning of GRK7, a candidate cone opsin kinase, from cone- and rod-dominant mammalian retinas." MOLECULAR VISION, vol. 4, no. 27, 8 December 1998 (1998-12-08), page NO PAGINATION XP002177859 ISSN: 1090-0535 the whole document	1-28
X	WO 95 07990 A (ICOS CORP) 23 March 1995 (1995-03-23) SEQ ID NO:12 the whole document	1-5
X	DATABASE EMBL 'Online! EBI; Accession No: AB002301, 1 July 1997 (1997-07-01) NAGASE T. ET AL.: "Human mRNA for KIAA0303 gene, partial cds." XP002177870 abstract	1-9
X	WO 94 18328 A (BIDEN TREVOR JOHN ;GARVAN INST MED RES (AU); SELBIE LISA (AU)) 18 August 1994 (1994-08-18) SEQ ID NO: 1 the whole document	1-28
X	DATABASE EMBL 'Online! EBI; Accession No: AC008735, 4 August 1999 (1999-08-04) DOE JOINT GENOME INSTITUTE STANFORD HUMAN GENOME CENTER: "Homo sapiens chromosome 19 clone CTD-253719" XP002177871 abstract	1-5
P,X	VERPLOEGEN SANDRA ET AL: "Identification and characterization of CKLiK, a novel granulocyte Ca++/calmodulin-dependent kinase." BLOOD, vol. 96, no. 9, 1 November 2000 (2000-11-01), pages 3215-3223, XP002177860 ISSN: 0006-4971 the whole document	1-28

-/--

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/32085

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>DATABASE EMBL 'Online! EBI; Accession NO: AK026945, 29 September 2000 (2000-09-29) SUGANO S. ET AL.: "Homo sapiens cDNA: FLJ23292 fis, clone HEP10334" XP002177872 abstract</p> <p style="text-align: center;">---</p>	1-5
X	<p>TOKUMITSU HIROSHI ET AL: "Characterization of a Ca-2+/calmodulin-dependent protein kinase cascade: Molecular cloning and expression of calcium/calmodulin-dependent protein kinase kinase." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 270, no. 33, 1995, pages 19320-19324, XP002177861 ISSN: 0021-9258 the whole document</p> <p style="text-align: center;">---</p>	1-28
X	<p>NAITO YASUHIRO ET AL: "Isoform-specific activation and structural diversity of calmodulin kinase I." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 272, no. 51, 19 December 1997 (1997-12-19), pages 32704-32708, XP002177862 ISSN: 0021-9258 the whole document</p> <p style="text-align: center;">---</p>	1-28
X	<p>HOCH BRIGITTE ET AL: "Identification and expression of delta-isoforms of the multifunctional Ca2+/calmodulin-dependent protein kinase in failing and nonfailing human myocardium." CIRCULATION RESEARCH, vol. 84, no. 6, 2 April 1999 (1999-04-02), pages 713-721, XP002177863 ISSN: 0009-7330</p> <p style="text-align: center;">---</p>	1-28
P,X	<p>DATABASE EMBL 'Online! EBI; Accession No: AC022132, 27 January 2000 (2000-01-27) DOE JOINT GENOME INSTITUTE: "Homo sapiens chromosome 5 clone CTD-2201J22" XP002177873 abstract</p> <p style="text-align: center;">---</p>	1-5
P,X	<p>WO 00 17232 A (REINARTZ JEANETTE ;BECHTEL MICHAEL (DE); KRAMER MICHAEL (DE); SCHA) 30 March 2000 (2000-03-30)</p> <p style="text-align: center;">---</p>	1-28
	-/--	

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/32085

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>WANG Z-N ET AL: "Cloning of a novel kinase (SIK) of the SNF1/AMPK family from high salt diet-treated rat adrenal" FEBS LETTERS, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 453, no. 1-2, 18 June 1999 (1999-06-18), pages 135-139, XP004259817 ISSN: 0014-5793</p>	1-28
P, X	<p>----- DATABASE EMBL 'Online!' EBI; Accession No: AF226044, 21 July 2000 (2000-07-21) LI Y. ET AL.: "Homo sapiens HSNFRK (HSNFRK) mRNA, complete cds" XP002177874 abstract</p>	1-28
X	<p>----- NAGASE T ET AL: "PREDICTION OF THE CODING SEQUENCES OF UNIDENTIFIED HUMAN GENES. III. THE CODING SEQUENCES OF 40 NEW GENES (KIAA0081-KIAA0120) DEDUCED BY ANALYSIS OF CDNA CLONES FROM HUMAN CELL LINE KG-1" DNA RESEARCH, UNIVERSAL ACADEMY PRESS, JP, vol. 2, no. 1, 28 February 1995 (1995-02-28), pages 37-43, XP001011422 ISSN: 1340-2838</p>	1-28
X	<p>----- BECKER WALTER ET AL: "Molecular cloning and characterization of a novel mammalian protein kinase harboring a homology domain that defines a subfamily of serine/threonine kinases." EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 235, no. 3, 1996, pages 736-743, XP002177864 ISSN: 0014-2956</p>	1-28
X	<p>----- HERRING B P ET AL: "DOMAIN CHARACTERIZATION OF RABBIT SKELETAL MUSCLE MYOSIN LIGHT CHAIN KINASE" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, vol. 265, no. 3, 25 January 1990 (1990-01-25), pages 1724-1730, XP000979556 ISSN: 0021-9258</p> <p style="text-align: center;">----- -/-</p>	1-28

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/32085

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>FISH KIMBERLY J ET AL: "Isolation and Characterization of Human Casein Kinase I-epsilon (CKI), a Novel Member of the CKI Gene Family."</p> <p>JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 270, no. 25, 1995, pages 14875-14883, XP002177865</p> <p>ISSN: 0021-9258</p> <p>figure 2</p> <p>---</p>	1-5
E	<p>WO 00 73469 A (PLOWMAN GREGORY D ;UGEN INC (US); WHYTE DAVID (US); MARTINEZ RICA) 7 December 2000 (2000-12-07)</p> <p>SEQ ID NO:46</p> <p>---</p>	1-28
P,X	<p>DATABASE EBI 'Online!'</p> <p>EMBL;</p> <p>Accession No: AF130372, 3 March 2000 (2000-03-03)</p> <p>MIDMER M., HAQ R., ZANKE B.W.: "NKIAMRE a novel kinase deleted in human leukemia"</p> <p>XP002177875</p> <p>abstract</p> <p>---</p>	1-28
X	<p>DATABASE EBI 'Online!'</p> <p>EMBL;</p> <p>Accession No: AA626859, 28 October 1998 (1998-10-28)</p> <p>HILLIER L. ET AL.: "WashU-NCI human EST Project"</p> <p>XP002177876</p> <p>abstract</p> <p>---</p>	1-28
X	<p>ABE M K ET AL: "EXTRACELLULAR SIGNAL-REGULATED KINASE 7 (ERK7), A NOVEL ERK WITH A C-TERMINAL DOMAIN THAT REGULATES ITS ACTIVITY, ITS CELLULAR LOCALIZATION, AND CELL GROWTH"</p> <p>MOLECULAR AND CELLULAR BIOLOGY, WASHINGTON, DC, US, vol. 19, no. 2, February 1999 (1999-02), pages 1301-1312, XP002172155</p> <p>ISSN: 0270-7306</p> <p>the whole document</p> <p>---</p>	1-28
X	<p>DATABASE EMBL 'Online!'</p> <p>EBI;</p> <p>Accession No: AW006971, 13 September 1999 (1999-09-13)</p> <p>NCI-CGAP: "National Cancer Institute, Cancer Genome Anatomy Project (CGAP)"</p> <p>XP002177877</p> <p>abstract</p> <p>---</p> <p style="text-align: center;">-/--</p>	1-5

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/32085

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL 'Online! EBI; Accession No: AF131745, 15 March 1999 (1999-03-15) MEI G. ET AL. : "Homo sapiens clone 25110 mRNA sequence, full-length" XP002177878 abstract</p> <p style="text-align: center;">---</p>	1-5
X	<p>ANAYA P ET AL: "Isolation of the Aspergillus nidulans sudD gene and its human homologue" GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES, ELSEVIER SCIENCE PUBLISHERS, BARKING, GB, vol. 211, no. 2, 12 May 1998 (1998-05-12), pages 323-329, XP004120595 ISSN: 0378-1119 the whole document</p> <p style="text-align: center;">---</p>	1-28
X	<p>TANAKA HIROMITSU ET AL: "Identification and characterization of a haploid germ cell-specific nuclear protein kinase (Haspin) in spermatid nuclei and its effects on somatic cells." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 274, no. 24, 11 June 1999 (1999-06-11), pages 17049-17057, XP002177866 ISSN: 0021-9258 figure 1</p> <p style="text-align: center;">---</p>	1-28
X	<p>DATABASE EMBL 'Online! EBI; Accession No: AC011189, 6 October 1999 (1999-10-06) BIRREN B., LINTON L., NUSBAUM C., LANDER E.: "Homo sapiens chromosome 17 clone RP11-231G16" XP002177879 abstract</p> <p style="text-align: center;">---</p>	1-5
X	<p>DATABASE EMBL 'Online! EBI; Accession No: AC010271, 17 September 1999 (1999-09-17) DOE JOINT GENOME INSTITUTE: "Homo sapiens chromosome 19 clone CTC-492K19" XP002177880 abstract</p> <p style="text-align: center;">---</p>	1-5
P,X	<p>US 6 001 623 A (BRUN KIMBERLY ANNE ET AL) 14 December 1999 (1999-12-14) the whole document</p> <p style="text-align: center;">---</p>	1-28

-/--



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/32085

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>MAUCUER A ET AL: "KIS is a protein kinase with an RNA recognition motif" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, vol. 272, no. 37, 12 September 1997 (1997-09-12), pages 23151-23156, XP002125937 ISSN: 0021-9258</p>	1-28
X	<p>BATTELLE BARBARA-ANNE ET AL: "A myosin III from Limulus eyes is a clock-regulated phosphoprotein." JOURNAL OF NEUROSCIENCE, vol. 18, no. 12, 15 June 1998 (1998-06-15), pages 4548-4559, XP002177867 ISSN: 0270-6474</p> <p>US 5 747 288 A (CHANG HENRY ET AL) 5 May 1998 (1998-05-05)</p>	1-28
X	<p>BIELKE W ET AL: "Characterization of a novel murine testis-specific serine/threonine kinase" GENE, ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, NL, vol. 139, no. 2, 1994, pages 235-239, XP002174517 ISSN: 0378-1119</p>	1-28
X	<p>DATABASE EMBL 'Online! EBI; Accession No: AC008476, 4 August 1999 (1999-08-04) DOE JOINT GENOME INSTITUTE STANFORD HUMAN GENOME CENTER: "Homo sapiens chromosome 5 clone CTC-380F16" XP002177881 abstract</p>	1-5
X	<p>DATABASE EMBL 'Online! EBI; Accession No: AB010154, 8 January 1999 (1999-01-08) NARA K.: "Cloning and characterization of a novel serine/threonine protein kinase gene expressed dominantly in developing brain" XP002177882 abstract</p>	1-28

-/--

# INTERNATIONAL SEARCH REPORT

National Application No  
PCT/US 00/32085

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL 'Online! EBI; Accession No: AC005374 4 August 1998 (1998-08-04) LEWIS S., WANG Y., ROE B.A.: "Homo sapiens Chromosome 22q11.2 Cosmid Clone cosi In NF1 Region" XP002177883 abstract</p>	1-5
X	<p>ZHENG CHAO-FENG ET AL: "Cloning and characterization of two distinct human extracellular signal-regulated kinase activator kinases, MEK1 and MEK2." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 268, no. 15, 1993, pages 11435-11439, XP002177868 ISSN: 0021-9258 the whole document</p>	1-28
X	<p>DATABASE EMBL 'Online! EBI; Accession No: AC011005 6 October 1999 (1999-10-06) KAUL R.K. ET AL.: "Homo sapiens chromosome 7 clone RP11-286H14" XP002177884 abstract</p>	1-5
X	<p>PARK S ET AL: "THE EEP RECEPTOR, A MEMBER OF THE EPH FAMILY OF TYROSINE PROTEIN KINASES, CAN BE ACTIVATED BY THREE DIFFERENT EPH FAMILY LIGANDS" ONCOGENE, BASINGSTOKE, HANTS, GB, vol. 14, no. 5, 6 February 1997 (1997-02-06), pages 533-542, XP001009718 ISSN: 0950-9232 the whole document</p>	1-28
X	<p>KLAUCK THERESA M ET AL: "Cloning and characterization of a glucocorticoid-induced diacylglycerol kinase." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 271, no. 33, 1996, pages 19781-19788, XP002177869 ISSN: 0021-9258</p>	1-28

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 00/32085

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
see FURTHER INFORMATION sheet PCT/ISA/210
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:  
1-28 (partially)
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-28 (all partially): Invention 1

An isolated kinase polypeptide comprising the amino acid sequence SEQ ID NO:58; other subject-matter referring to said polypeptide.

2. Claims: 1-28 (all partially): Inventions 2-56

Idem as subject 1 but limited to each of the polypeptides comprising the amino acid sequences SEQ ID NO:59-114

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 15-20 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claims 21-28 cover a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

I. International Application No

PCT/US 00/32085

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9507990 A	23-03-1995	US 5532151 A US 5591618 A	02-07-1996 07-01-1997
WO 9418328 A	18-08-1994	AU 673897 B AU 5996394 A EP 0644939 A JP 8500018 T US 5595902 A	28-11-1996 29-08-1994 29-03-1995 09-01-1996 21-01-1997
WO 0017232 A	30-03-2000	DE 19842863 A AU 1149100 A EP 1114157 A	27-04-2000 10-04-2000 11-07-2001
WO 0073469 A	07-12-2000	AU 5173400 A	18-12-2000
US 6001623 A	14-12-1999	EP 1100882 A WO 0006709 A	23-05-2001 10-02-2000
US 5747288 A	05-05-1998	US 5700675 A AU 1291097 A EP 0870035 A JP 2000501943 T WO 9721820 A US 5747275 A	23-12-1997 03-07-1997 14-10-1998 22-02-2000 19-06-1997 05-05-1998

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**BEST AVAILABLE COPY**